DOCTORAL THESIS

The Perception of Transformed Auditory and Visual Pattern Structure: An Exploration of Supramodal Pattern Space

Thorpe, Michael J A

Award date:
2016

Awarding institution:
University of Roehampton

The Perception of Transformed Auditory and Visual Pattern Structure: An Exploration of Supramodal Pattern Space

by

Michael J A Thorpe BSc, MSc

A thesis submitted in partial fulfilment of the requirements for the degree of PhD

Department of Psychology

University of Roehampton

2015

Abstract

The present thesis is broadly concerned with the processing of structural information. More specifically, it investigates the possibility that auditory pitch patterns share, at some level, supramodal structural representations and processes with visuo-spatial patterns. The motivation for the research was provided by a number of areas of psychological research that are brought together and discussed in this thesis, and which inform the development of a new theoretical framework that conceives of a supramodal pattern space (SPS). According to the SPS framework, auditory and visual patterns can be represented in equivalent '1½-D' supramodal pattern spaces. A series of experiments was devised to test the assumptions of the SPS framework, by means of analysing the perception of two types of structural transformation: inverse and retrograde. The main hypothesis that was tested in all experiments predicted a processing advantage for inverse transformations when patterns corresponded to 1½-D supramodal pattern space. Support for the hypothesis was provided by experiments adopting a short-term recognition paradigm. However, contrasting results were revealed by experiments adopting a structural priming paradigm, which did not support the hypothesis. It was concluded that different processing strategies were used depending on the task demands. The findings were discussed with relation to theories of sequential pattern learning, melodic perception and brain organisation.

Table of Contents

List of figures

List of tables

Acknowledgements

Firstly, I would like to thank my supervisors – Dr Aleksandar Aksentijević, Professor Michael Eysenck, and Professor Adam Ockelford – for their guidance, advice, and constructive criticism. In particular I would like to thank Dr Aleksandar Aksentijević for his enthusiasm for the research, and for the many inspirational discussions.

Secondly, I would like to thank the Department of Psychology for making this project possible and for providing funding. Special thanks goes to Martin Evans who helped me with many technical issues, which included sourcing and programming the software needed for conducting one of my experiments. Thanks also to the fantastic community of research students at Whitelands College, of which I have been fortunate to be a member.

Finally, I would like to express my deepest gratitude to my family and friends for their enduring support and patience during what has been a challenging endeavour.

Chapter 1: Literature review

1.1 Introduction to the thesis

The research reported in the present thesis is concerned with the possibility that auditory pitch patterns share, at some level, supramodal structural representations and processes with visuo-spatial patterns.

The motivation for the research is provided by a number of areas of psychological research that are discussed in some detail in the present chapter. Firstly, strikingly similar (if not the same) principles appear to govern pattern perception in both the auditory and the visual domains. Secondly, there is growing evidence for the spatial representation of psychological dimensions such as auditory pitch and time. Thirdly, neuropsychological studies have shown that auditory pitch patterns and visuo-spatial patterns may be processed in shared higher-order anatomical areas of the cortex. Specifically, areas of the posterior parietal cortex have been associated with the processing of melodic transformations and visuo-spatial transformations.

The main aim of the research is to explore possible supramodal processes more thoroughly and in more detail than has been achieved before now, by means of behavioural experimentation. To this end, a theoretical framework is proposed in Chapter 2 that conceives of a supramodal pattern space (SPS). According to the SPS framework, structural information, abstracted from sensory information, can be represented on one or a combination of qualitatively distinct supramodal dimensions. Two such dimensions are identified as being required to represent simple auditory pitch patterns (monophonic, atonal melodies): 1) a scalar dimension, which represents relative pitch, and 2) a temporal dimension, which represents the relative timing of auditory events. This supramodal pattern space is labelled a 1½-D space to reflect the qualitative distinction between the dimensions from which it is constructed, in terms of their directionality – the scalar dimension is bidirectional because events can move along it in either direction, and for this reason it is considered to be a whole spatial dimension; the temporal dimension is unidirectional, reflecting the fact that experienced time unfolds in one direction only, and for this reason it is considered to be half a spatial dimension.

The SPS framework provides a means of comparing the processing of equivalent (or non-equivalent) structural information presented in different sensory modalities (auditory and visual), and was tested by examining the perception of pattern regularities described by two special types of isomorphic structural transformation: *inverse* and *retrograde*. For patterns represented in a 1½-D supramodal pattern space, the perception of inverse transformations requires an inversion of ordinal relations on the scalar dimension, whilst the perception of retrograde transformations requires an inversion of ordinal relations on the temporal dimension. One of the assumptions of the SPS framework is that inversions on the temporal dimension would be harder to process, due to the dimension's inherent directionality. Thus, the SPS framework provided a testable hypothesis: when stimuli correspond to representations in a 1½-D supramodal pattern space, structural regularities should be perceived more effectively when they are described by inverse compared to retrograde transformation, irrespective of the sensory modality from which structural information is abstracted.

Chapter 3 of the thesis reports a preliminary experiment (Experiment 1), which was carried out to investigate the possibility that auditory pitch space and visual space share a common metric. This possibility has received little (or no) attention in previous experiments that have also made a structural analogy between auditory pitch and visuo-spatial patterns. Analysis of the results provided a pitch-interval to visual-distance ratio that was used when designing the stimuli used in all subsequent experiments.

Chapters 4, 5 and 6 report the main body of the research, which tested the 1½-D hypothesis, described above. The paradigm adopted for the 5 experiments reported in Chapters 4 and 5 was a short-term recognition paradigm, which required participants to identify when target patterns were a transformation (inverse, retrograde) of a preceding standard pattern, and when they were not. Chapter 4 reports Experiments 2 and 3, which investigated the perception of structural transformations when auditory and visual stimuli corresponded to non-equivalent supramodal pattern spaces. Chapter 5 is a continuation of Chapter 4, and reports three more experiments (Experiments 4, 5 and 6) investigating the perception of structural transformations when the stimuli corresponded to an equivalent supramodal pattern space.

The experiments reported in Chapter 6 (Experiments 7 and 8) employed an alternative structural priming paradigm, to see if transformations would be processed when participants had neither been explicitly instructed to compare patterns, nor had been informed of the way in which the patterns they encountered might be related. An additional aim of these experiments was to explore the time courses of hypothesised supramodal mechanisms.

Finally, Chapter 7 provides a summary and general discussion of the research.

1.2 Outline of the chapter

The following sections of the present chapter will introduce the theoretical basis and rationale for the present thesis, drawing upon scientific literature from a number of areas of psychological research including cognitive psychology, music psychology, psychophysics and neurocognitive psychology.

Section 1.3 introduces the melodic transformations that are the focus of the current research. Section 1.4 discusses theoretical and empirical literature relating to the perception and cognition of pattern structure. Perception is held to be governed by a *simplicity principle* that promotes the simplest interpretations of structural information. It is argued that simple interpretations are important because they minimise processing cost, and that the simplest interpretations are achieved by detecting structural regularities within and between patterns. Section 1.5 reviews research addressing the perception of global structural regularities in static visual patterns, and in sequential patterns. The transformational approach to formalising pattern structure is introduced, which views regularities as aspects of a shape or pattern that remain invariant under certain transformations (Palmer, 1983). Next, melodic processing literature is discussed, highlighting the importance of relative over absolute information in pattern perception, and the role of ordinal (i.e. contour) and interval structure.

In Section 1.6, experiments demonstrating the spatial representation of pitch and time are reported. It is proposed that these reflect partly shared representations and processes for auditory pitch patterns and visuo-spatial patterns. Section 1.7 pursues the idea that global pattern structure, abstracted from sense-specific information (the focus is on auditory and visual information), is processed by shared mechanisms in the brain by seeking the potential neural correlates, highlighting areas in the parietal cortex. Finally, the orientation of the thesis is summarised in Section 1.8.

1.3 Introduction to melodic transformations

In music, a melody may undergo special kinds of symmetry-producing structural transformation, displayed in Figure 1.1. A *retrograde* transformation involves a reversal in the temporal order of a melody's notes, effectively producing a backwards version of the original. An *inverse* transformation maintains the temporal order of a melody's notes, but inverts the direction of the pitch changes between them – any upwards pitch change becomes a downwards pitch change, and any downwards pitch change becomes an upwards pitch change. Both of these can be combined into a *retrograde inverse* transformation – the temporal order of notes is reversed and the direction of pitch changes between them is inverted.[1]

These types of melodic transformation can be found in many forms of music, and are particularly notable for their use as a compositional technique in Western classical styles. An early example of retrograde transformation may be traced back to the 13th century (found in the manuscript Pluteo 29.1, folio 150 verso, located in the Laurentian Library in Florence), and it receives mention in the music theoretic literature from as early as the 16th century onwards (e.g. see writings by Nicola Vicentino, 1555, and Thomas Morley, 1597). More recent examples may be found in the works of Joseph Haydn (e.g. canon "Thy Voice, O Harmony"; Symphony no. 47, 3rd movement, "Minuetto al Roverso"; piano

[1] Inverse and retrograde transformations can be combined to make either retrograde inverse transformations or inverse retrograde transformations, depending on the order in which they are applied. Essentially, both types of combined transformation produce the same melody. In the music theory and music psychology literature, the term 'retrograde inverse' is used more commonly than 'inverse retrograde'. In the present thesis, retrograde inverse is used to refer to a combination of inverse and retrograde transformations in either order.

sonata, XVI/26, minuet [a transcription of the 3rd movement from Symphony no. 47]; Violin Sonata no. 4). Examples of inverse transformations may be found in Johann Bach's Two and Three Part Inventions in C Major and, though less utilised, a good example of a retrograde inverse transformation may be heard in Igor Stravinsky’s Requiem Canticles. All three types of transformation can be found in the “serialist” music of composers adopting the twelve-tone technique that was first developed by Arnold Schoenberg in the early part of the 20th century (e.g. Anton Webern, Karlheinz Stockhausen and Pierre Boulez). Integral to this technique is the use of tone rows, composed from the 12 notes of the chromatic scale, that are subjected to the three types of transformation described above (see Schoenberg, 1975).

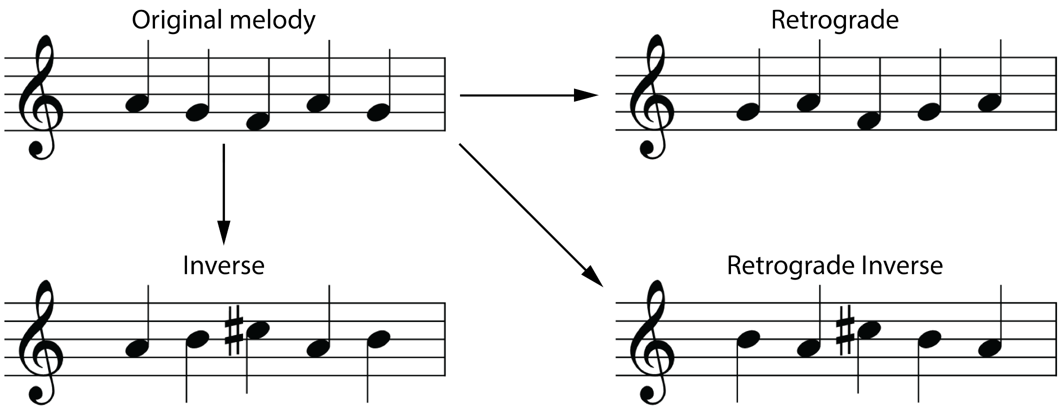


Figure 1.1. Example melody and its transformations. Here the transformations are exact because they preserve the size of the original melody’s pitch intervals. Transformations can also be applied that adjust the pitch interval sizes to conform to a particular diatonic key. Furthermore, the transformations displayed here are untransposed (the melody under retrograde transformation begins on the note that the original melody ends on; the melody under inverse transformation begins on the same note as the original melody), but they can also be transposed to begin on any other note.

Although there is some debate regarding their perceptual validity, with some arguing that they are nothing more than an interesting compositional technique (Levarie & Levy, 1981; Mazzola, 2013; Morgan, 1998), a number of psychological investigations have demonstrated that listeners can (to a certain extent) recognise these melodic transformations (Cupchik, Phillips, & Hill, 2001; Dowling, 1972; Krumhansl, Sandell, & Sergeant, 1987; Schulze, Dowling, & Tillmann, 2012; White, 1960). In fact, the psychological relevance of these structural transformations transcends music, and even the auditory domain. Psychologists interested in the processing of sequential information more generally have shown that inverse and retrograde transformations describe two structural regularities that can be used by human observers to represent patterned sequences of letters and numbers (Jones & Zamostny, 1975; Kotovsky & Simon, 1973; Leeuwenberg, 1969; Vitz & Todd, 1969), or visual objects distributed in space (Fountain & Rowan, 1995; Koch & Hoffmann, 2000; Kundey et al., 2013; Restle & Brown, 1970; Restle, 1970). More than this, it has been proposed that inverse and retrograde transformations belong to a limited set of regularities that, in isolation or in combination with others, can describe the structure of any possible sequential pattern that can be encountered and perceived by an organism (Jones, 1974, 1978). Thus, these melodic transformations are more than just musical curiosities – they have clear analogues in visual processing which have not been sufficiently investigated. Studying them provides a potentially important window into how the brain deals with pattern.

1.4 Pattern perception: some theoretical considerations

Before addressing the processing of particular structural regularities, it is necessary to discuss pattern processing more generally. The processing of pattern may be considered one of the fundamental functions of the human perceptual system. According to the Oxford English Dictionary, one meaning of the word 'pattern' is "*a regular and intelligible form or sequence discernible in the way in which something happens or is done*". This definition reflects the rather abstract conceptual meaning of the word, which is difficult to convey in a single sentence. It can be broken down to illustrate three important features of a pattern (in context of the present work): 1) a pattern is *regular* – it is something that follows rules and may involve repetition; 2) a pattern is *intelligible* – a pattern must be simple enough to be perceived by an organism for it to be of any subjective relevance; 3) a pattern is a *form* or a *sequence* – it consists of parts that relate to each other and make a whole either in space or in time.

Humans receive far more information from the environment than they could possibly process. A key survival strategy that has evolved phylogenetically and develops ontogenetically is the ability to make assumptions relating to incoming information that permit it to be processed, stored and retrieved parsimoniously. Crucial to this is pattern. In the natural environment, patterns abound, and can be found in both inorganic form (e.g. the ebb and flow of ocean waves, the strata of rocks, the six-fold symmetry of snowflakes) and in organic entities (e.g. the bilateral symmetry of mammals, the fractal growth of trees, Fibonacci spirals in the head of the sunflower). Patterns also abound in the world of human creation and may be found in the artworks and architecture of ancient and modern societies from all over the globe (see Weyl, 1952, for an in-depth discussion). Humans are highly sensitive to patterns and rely on pattern perception to perform all manner of everyday tasks, from recognising faces to reading the newspaper and listening to music. Furthermore, our ability to perceive and recognise patterns can be utilised to achieve far greater accomplishments such as predicting the dates of solar eclipses, or formulating the evolutionary theory of natural selection. Beyond this, humans and other organisms are seemingly biologically predisposed to entrain to environmental rhythmic patterns of daily, seasonal and yearly change (e.g. circadian rhythms, see Koukkari & Sothern, 2006).

Pattern perception probably involves a process that is more than simply the perception of isolated features, component parts or objects. In order for a pattern to be perceived, the relationship between all of these elements must also be processed – in other words, pattern perception must involve the processing of *structure*. The perception of structure is quite a different matter to the perception of physical stimuli, as structure is an abstract property that transcends any particular stimulus (Pomerantz & Lockhead, 1991, p.5). The importance of structure in perception was emphasised by the Gestalt psychologists of the early to mid-20th century (Koffka, 1935; Köhler, 1929; Wertheimer, 1912, 1922, 1923), who proposed that structured wholes – or *Gestalten* – make the primary units of mental life.[2] In their view, the perceptual system considers the global whole in parallel with its local parts. What appears as a whole and what appears as a part is determined by the functional relations between them – the whole is qualitatively

---

[2] The Gestalt approach contrasted with the preceding focus on psychophysics which was more concerned with the relationship between the physical attributes of stimuli, such as quantity and magnitude and sensation (see Ernst Weber and Gustav Fechner).

different from what one might predict by considering only its parts, and the quality of a part depends upon the whole in which it is embedded.

Although Gestalt theory was established as a general theory of perception, early thinkers were particularly interested in auditory processes, and even used musical examples to illustrate their ideas (e.g. von Ehrenfels, 1890, 1937). Nevertheless, over recent years Gestalt principles have been applied more frequently to the visual domain. This bias towards vision perhaps started with Wertheimer (1923) who proposed a number of principles that describe the functional relations leading to the emergence of Gestalten, using visual examples to illustrate them. He demonstrated that perceptual grouping in vision may be described by principles such as proximity, similarity, good continuation, common fate, closure and symmetry (these were later developed by Koffka, 1935). These principles have proved to be highly influential in the study of perception, and subsequent research in vision has succeeded in quantifying the strength of certain grouping principles in both static patterns (Hochberg & Silverstein, 1956; Quinlan & Wilton, 1998) and dynamic ones (Burt & Sperling, 1981; Wallace & Scott-Samuel, 2007). In addition to these classic grouping principles, new principles (or extensions of existing principles) continue to be identified and investigated (for a recent review, see Wagemans et al., 2012). Despite the focus on visual perceptual organisation, many grouping principles have also since been found to affect perception in other sensory modalities, such as touch (Gallace & Spence, 2011) and audition (Bregman, 1978, 1990). This has led to the suggestion that the dynamics of perceptual organisation may be associated with common principles that can be described mathematically (Aksentijevic, Elliott, & Barber, 2001).

Importantly, Wertheimer (1923) also proposed a general law, the law of Prägnanz, that underpins all human perception and which states that of all the possible ways of interpreting the perceptual field and the objects within it, the simplest and most encompassing will be selected.[3] One of its fundamental implications is that the perceptual scene is organised in such a way as to minimise the expenditure (or rather conversion) of energy. From this perspective, Koffka (1935) conceived perception as a neural system that, when presented with a stimulus, exhibits the tendency to settle into an equilibrium involving minimum energy load; the resulting neural pattern of activation then forms the mental representation of the stimulus. Associated with the law of Prägnanz is the notion of pattern "goodness", which refers to a pattern's salience or perceptual strength. Generally, patterns are considered "good" when they are structured, simple, orderly and regular. For example, the Gestaltists assigned a high level of goodness to patterns that contain mirror symmetry (Koffka, 1935). In contrast, patterns are considered "poor" when they are unstructured, complex, disorderly and irregular. Thus, the law of Prägnanz proposes that the perceptual system is driven towards "good" organisations of sensory information, which minimise energy cost.

The Gestalt law of Prägnanz has since been interpreted, by way of Shannon's (1948) information theory, as the *simplicity principle* (Chater, 1996;

[3] This idea can be traced back to Occam's Razor (William of Occam, circa. 1290-1349) which advised that, all else being equal, the simplest of all possible interpretations of data is the best one.

Hochberg & McAlister, 1953).[4] According to Hochberg and McAlister's (1953) interpretation, the simplicity principle holds that "the less the amount of information needed to define a given organisation as compared to the other alternatives, the more likely that that interpretation will be perceived" (p.361). They specify descriptive information loads, or complexities, as the number of different units of information needed in order to specify or reproduce a given pattern. This specification has effectively given rise to the modern notion of *pattern complexity* (Aksentijevic & Gibson, 2012a; van der Helm, 2000). In general terms, the complexity of a pattern is defined as the minimum amount of information needed to describe it. Around the same time as Hochberg and McAlister (1953) published their interpretation of the simplicity principle, Attneave (1954) demonstrated how the perceptual system might arrive at the simplest interpretation of a pattern by utilising the information-theoretic concept of *redundancy*. Redundancy was originally conceived as a measure of the difference in the amount of information that a source is capable of emitting (i.e. its *entropy*), and that of the message (Shannon, 1948). If the successive symbols of a message are not sent with equal probability, but follow certain rules, a symbol may be partly predictable from what has come before it. As a result, the message contains some structure that does not convey any additional information. The more structured a message, the more redundant it is. In a psychological context,

[4] It should be noted that the simplicity principle contrasts directly with von Helmholtz's (1909/1962) likelihood principle, which views perceptual organisation as being guided by veridical knowledge which yields the most likely interpretation, based on the probabilities of previous experience. The simplicity principle and the likelihood principle may make the same predictions (the most likely interpretation is often the simplest) and there is continued debate over which principle may give a better explanation for perceptual interpretations in different contexts (Chater, 1996; van der Helm, 2000).

redundancy refers to the proportion of perceptually "excessive" information in a pattern – excessive because it can be predicted from existing information.

Inspired by this information-theoretic approach, a number of coding theories have been developed to model the way in which structured information is processed (for an early review see Simon, 1972). The most highly developed of these coding theories are concerned with visual perception, and have focussed on regularities in static patterns (e.g. algorithmic information theory [AIT]: Kolmogorov, 1968; Li & Vitányi, 1997; Solomonoff, 1964; Vitányi & Li, 2000; and structural information theory [SIT]: Buffart, Leeuwenberg, & Restle, 1981; Leeuwenberg, 1969; van der Helm & Leeuwenberg, 1991, 1996; van der Helm, 2000). Collectively, they assume that humans abstract the simplest possible representation from a stimulus (in line with the simplicity principle; Hochberg & McAlister, 1953), and that the simplest representations are obtained by using a systematic relationship or set of relationships among rules to relate pattern elements. To model this process, a stimulus is described by a code string, which can be compressed according to structural regularities (i.e. rules) that are present in the stimulus. The complexity of a pattern is equal to the length of the shortest statement that can encode it. Crucially, these theories emphasise the importance of pattern perception as a way of signalling redundancy to the observer and propose a framework for how redundant information might be used to arrive at the simplest interpretations.

Despite their influence in psychology, theories of complexity based on information theory and coding theory have been criticised for a number of reasons (see Aksentijevic & Gibson, 2012a). Most fundamentally, it is argued that the issue of information and its cost has not been properly addressed. As noted earlier, one of the implications of the simplicity principle is that the perceptual scene is organised in such a way as to minimise the expenditure (or rather conversion) of energy. Measures of complexity that involve coding algorithms cannot properly explain how arriving at the simplest possible description can minimise energy expenditure – arriving at the simplest possible code for a simple pattern can require more complex computations than arriving at a code for a more complex pattern which is relatively incompressible. However, it is clear that simple patterns involve less effort and energy to process than complex patterns (e.g. Falk & Konold, 1997).

Aksentijevic and Gibson (2012a) have proposed an alternative approach to psychological complexity that is founded on the primitive notion of *change*. They explain that any perception, cognition or action involves change, and is accompanied by an irreversible expenditure of energy. Change equals an increase in energy conversion, and this in turn equals cost. The notion of change as a measure of complexity is particularly attractive because it focuses attention on the relationship *between* elements of a pattern rather than on the elements themselves, thereby capturing structural complexity. To test this theory the authors quantified an index of structural complexity and compared it with a host of existing complexity measures (Aksentijevic & Gibson, 2012b). It correlated highly with both objective and subjective measures of complexity addressing both visual and auditory domains. The authors concluded that this provides evidence that change represents the ‘conceptual core of complexity’ (Aksentijevic & Gibson, 2012a, p. 14).

The importance of structure has been acknowledged by coding theorists. For example, structural information theory (SIT: van der Helm, 2000) differentiates between two types of information that appear to be used with unequal effectiveness by humans: *metrical*, which refers to precise numerical values of information (e.g. describes the exact sizes of pattern elements, such as the length of a line segment in a square), and *structural*, which describes the relationship between pattern elements (e.g. describes the fact that all four sides of a square are the same size).[5] However, their theories have modelled pattern perception as a bottom-up process that begins with a description of a pattern's metrical information and applies structural rules to compress the code post-hoc. This approach is not consistent with the Gestalt view that the perceptual system considers the global whole in parallel with its local parts (Koffka, 1935; Köhler, 1929; Wertheimer, 1912, 1922, 1923). By taking into account the importance of the processing cost associated with change, it becomes apparent that the processing of structure is of fundamental importance because it is structure that signals redundancy, which can be used to predict pattern elements and save processing cost. Furthermore, it also becomes apparent that the processing of higher order regularities that encompass the largest number of pattern elements are arguably of the greatest importance in pattern perception: these global regularities signal the greatest proportion of redundant information and therefore allow the greatest saving of energy.

1.5 The processing of global structural regularities

The melodic transformations introduced in Section 1.3 are examples of global structural regularities. A global structural regularity is defined here as one

[5] This distinction was based on MacKay's (1950) decomposition of the classical concept of information into two concepts of information: the *metron* and the *logon*.

which relates groups of pattern components. A local structural regularity, on the other hand, is one which relates individual components of a pattern. The processing of global structural regularities has been well researched in the visual domain, though the focus has been on static patterns, or objects, as opposed to dynamic, sequential patterns that unfold over time. This does not mean that work in this area is not relevant to the present research. Recently, the notion of objecthood has been re-examined and researchers are beginning to focus on the similarities between modalities rather than on the differences (e.g. Bizley & Cohen, 2014; Griffiths & Warren, 2004; Kubovy & Valkenburg, 2001; Shamma, 2008; Turatto, Mazza, & Umiltà, 2005).

According to Bizley and Cohen (2014) an auditory object is a perceptual construct that corresponds to the sound that can be assigned to a particular source. It is the result of the auditory system's ability to detect, extract, segregate and group the spectrotemporal regularities in the acoustic environment into stable perceptual units. Auditory objects have a number of characteristics and features that are strikingly comparable to those of visual objects – two of which are of particular relevance. First, visual and auditory objects can be hierarchically formed from other objects. An object can be decomposed into component parts that, in isolation, can also be considered to be objects. Examples in the visual domain include a square that is composed of four sides, or a tree that is composed of leaves, branches, a trunk and roots. Examples in the auditory domain include a single note played on a musical instrument that is composed of multiple harmonic frequencies, or indeed a melody that is composed of multiple tones of different pitch.

Second, visual and auditory object recognition is invariant to changes in the absolute properties of the stimulus. Absolute properties correspond to precise measurements, such as the length of the sides in a square and their placement in space, or the frequency and amplitude of the harmonic components in a tone. An object may retain its identity when these absolute properties are changed, provided they preserve their relational properties – the sides of a square can change size and spatial position, but must be of the same proportion and position with relation to each other; the tones of a melody can change pitch, but their fundamental frequencies must stand in the same ratio (the absolute and relative properties of melodic stimuli are discussed in more detail in Section 1.5.3).

On the assumption that visual and auditory objects are in some way theoretically equivalent, first some relevant research in visual symmetry and similarity perception is discussed; second, some classic research on the topic of serial pattern learning is discussed: although rooted in coding theory (which has been criticised above), this research has specifically addressed the processing of sequential pattern structure; third, research addressing melodic processing is discussed, and ultimately, some behavioural research investigating the perception of inverse and retrograde transformations of melodic structure is reviewed.

1.5.1 Visual pattern perception: symmetry and similarity

In order for the perceptual system to utilise redundant information and minimise processing cost, it must detect regularities. In the visual domain, the detection of regularity has been well researched. A distinction is made between *intra*-pattern regularity (i.e. structure within a single object) and *inter*-pattern regularity (i.e. structure between two or more objects), though there is some crossover in terms of the theoretical discussion that addresses why certain regularities are perceptually relevant (e.g. Chater, 1999). In the case of static patterns, pattern structure has typically been formalised in terms of a transformational approach (TA), according to which, regularities represent those aspects of a shape or pattern that remain invariant under certain transformations (Palmer, 1983). This approach to defining pattern regularity in perception evolved from the application of mathematical principles to the formalisation of structure in nature (Weyl, 1952) and considers pattern structure to consist of three types of regularity that can exist in isolation or in combination: translation, rotation, or reflection (see Figure 1.2).

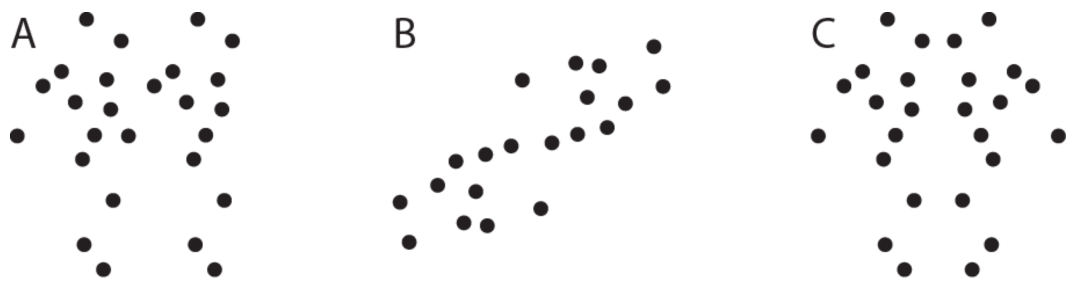


Figure 1.2. Examples of visual symmetry (adapted from Wagemans, 1997). Random-dot patterns with (A) translational symmetry, (B) rotational symmetry and (C) reflectional symmetry.

Empirical research in the field of visual symmetry perception has confirmed the perceptual validity of intra-pattern regularities described by these transformations (for reviews see Treder, 2010; Wagemans, 1995). However, this approach has not been completely successful in explaining why certain visual regularities are ‘better’ than others (van der Helm & Leeuwenberg, 1996). TA implies that perceptual representations of regularities are given a ‘block structure’. In other words, the structured whole of a symmetric pattern is represented by 'blocks' of sub-structures (which represent sub-patterns) that can be mapped onto each other. For the perceptual system to detect regularities in this way, it requires the perceptual representation and manipulation of stimuli to be analogous to physical stimuli. Accordingly, the transformation process required to detect invariance has been attributed to mental rotation (see Shepard & Metzler, 1971; Tarr & Pinker, 1989). This predicts that one-fold mirror symmetry (described by reflectional transformation), repetition (described by translational transformation) and rotational symmetry are equally redundant (because in each case their detection would involve a single transform operation) and are therefore equally "good". But this is not the case – it is generally found that mirror symmetry is more salient than rotational symmetry (Julesz, 1971; Palmer & Hemenway, 1978; Royer, 1981).

Attneave (1954) suggested that mirror symmetric patterns must possess an extra kind of redundancy that is distinct from the block structure implied by TA. He elaborated that mirror symmetric patterns could be described by the relationship of each point in the figure to a single axis of symmetry, implying that its perceptual representation may be given a 'point structure'. A point structure representation would not require a transformation to detect the regularity, and would explain why mirror symmetry is perceived more quickly than other regularities that involve a transformation of some kind. In response to issues such as those presented above, the holographic approach (HA), developed within the framework of SIT (van der Helm & Leeuwenberg, 1996), offers an alternative approach to the formalisation of visual regularities by extending TA to include Attneave's description of a point structure for mirror symmetry. However, it should be noted that TA and HA are just two of many theoretical models that have been proposed to account for the human perception of regularities, and none have so far been able to give a comprehensive account as they are too stimulus specific (for a review see Treder, 2010). Following the focus of the present research, it may be suggested that a more successful account of visual symmetry may be found by emphasising the importance of structural representations over representations that are analogous to visual stimuli, and therefore stimulus specific.

A transformational approach has also been used to address the problem of finding inter-pattern regularities, commonly referred to as similarity perception. Similarity is defined as the degree of resemblance between two objects or events (Hahn, 2014, p.1). Traditional theoretical accounts of similarity are the spatial account (Shepard, 1957) and the featural account (Tversky, 1977). Shepard's (1957) spatial account represents objects as points in an internal psychological space. An object's position is determined through its coordinate values along the relevant psychological dimensions, and the similarity between two objects is inversely related to the distance between their representations in this space. Tversky's (1977) featural account, on the other hand, views similarity as a function of common and distinctive features of an object's entities under comparison. This account successfully addressed problems faced by the spatial account, such as the assumption of symmetry (the spatial account predicts that the similarity relationship between two distinct objects is symmetrical, i.e. the same regardless of the direction in which they are compared, which is not always the case).

Although the spatial and featural accounts of similarity have both received empirical support (Nosofsky, 1986; Ortony, 1979; Shepard, 1987; Tversky, 1977), they have been criticised for not addressing the importance of structural relationships within and between objects (Hahn, 2014). As discussed above, psychological representations of complex stimuli are assumed to be structured, so that the whole can be decomposed into component parts, the part into subparts, and so on (Koffka, 1935). Objects represented as points in space or as a list of features cannot capture structured descriptions of objects. A more recent account, called *representational distortion* (RD) (Hahn, Chater, & Richardson, 2003), has incorporated the importance of structural representations to propose a general framework for understanding similarity. According to this interpretation, if two objects are highly interrelated and share patterns, they are considered to be similar. The fewer patterns shared by distinct objects, the less similar they are judged to be. This is a transformational account that views the similarity between two object representations as a function of the complexity of the transformation that is required to produce one representation from the other (i.e. the length of the shortest algorithm that transforms or distorts one representation into the other). The simpler the transformation that the cognitive system is able to find between the representations of two objects, the more similar those objects are assumed to be. The notion of similarity can therefore be equated with the Gestalt concept of pattern goodness. This account has clear affinities with approaches to understanding intra-pattern symmetry perception outlined earlier. However, the detection of intra-pattern regularity may not be directly equivalent to the detection of inter-pattern regularity – the perception of certain symmetries appear to proceed differently depending on whether the symmetry occurs within a pattern or between two perceptually distinct objects (Olivers & van der Helm, 1998).

Hahn et al. (2003) carried out three experiments in which simultaneously presented pattern pairs were rated for similarity. Stimuli consisted of binary sequences of filled and unfilled circles, simple geometric shapes, or arrangements of Lego bricks. In each pair, one of the patterns had undergone either a single transformation or a combination of up to six different types of transformation. The types of transformation used included those identified as perceptually relevant in the intra-pattern regularity literature such as mirror and rotational symmetry. Also included were transformations such as insertion, deletion, phase shift, stretching of the whole object, and spatial rearrangements of the pattern components. In all experiments, a statistical relationship was found between transformation distance (the number of transformations applied to one object) and perceived similarity, offering support for the transformational account of pattern similarity. More recently, transformational similarity was also found to predict reaction time in speeded same-different judgements of sequentially presented shape pairs (Hodgetts & Hahn, 2012).

1.5.2 Sequential pattern perception

At around the same time that coding theories were being developed to address visual perception, they were also being developed to address the perception of sequential patterns (often by the same researchers, e.g. visual perception: Buffart et al., 1981; Leeuwenberg, 1968; sequential pattern perception: Leeuwenberg, 1969; Restle, 1970). The focus of these theories was initially on the problem of serial pattern learning – theorists sought to explain how humans are able to learn complex sequences via the internal representation of pattern regularities. Unlike theories applied to the perception of static visual patterns, they were not specific to any particular sensory modality, and early empirical work investigating these theories was carried out using a variety of stimuli such as letter and number series (Jones & Zamostny, 1975; Kotovsky & Simon, 1973; Leeuwenberg, 1969; Vitz & Todd, 1969), spatial patterns of lights (Restle & Brown, 1970; Restle, 1970, 1976) and musical patterns (Collard & Povel, 1982; Deutsch & Feroe, 1981; Deutsch, 1980). In general, this work has developed the idea that when humans encounter patterned sequences, they generate an abstract representation (i.e. rule) of the sequence's structure that describes how the events are organised within the sequence, which can then be used to generate the entire sequence. Importantly, the individual events in a sequence are not encoded in their entirety but as subsequences, or "chunks", that allow information to be compressed into a form that lessens memory demands. A variety of these coding models have been proposed (e.g. Garner & Gottwald, 1968; Jones & Zamostny, 1975; Leeuwenberg, 1969; Simon & Kotovsky, 1963; Vitz & Todd, 1969), but the most influential remains the hierarchical model (Restle & Brown, 1970; Restle, 1970).

The hierarchical model assumes that the simplest possible representation is abstracted from a sequence by using lower-order rules that relate individual pattern elements to create subsequences, as well as higher-order rules that relate subsequences of information. Thus, a nested, hierarchical organisation can be constructed in which the highest-order rules relate the largest number of pattern elements, while the lower-order rules are nested within the higher-order structure. Within Restle's coding system, rules are applied to a given alphabet (e.g. 1 2 3 4 5 6). Restle suggested that any operation rule can be used, provided the relationship it describes is psychologically valid (i.e. can be used by the observer). Three rules were originally proposed: repeat, transpose and mirror image (Restle & Brown, 1970; Restle, 1970), but other rules were subsequently added, including temporal inversion (Restle, 1976). The repeat rule is self-explanatory. The transpose rule shifts an event a specified number of steps along the alphabet scale. The mirror image rule involves a complementary transform of, for example, 1 into 6, 2 into 5, and so on, whilst the temporal inversion rule is applied to a minimum of two events and reverses their temporal order (e.g. 12 becomes 21). The mirror image and temporal inversion rules can be considered equivalent to the inverse and retrograde transformations of pattern structure that are the focus of the experiments reported in the present thesis. Restle combined these rules to create patterns with hierarchical structure that could be illustrated as structural trees. Two examples are displayed in Figure 1.3.

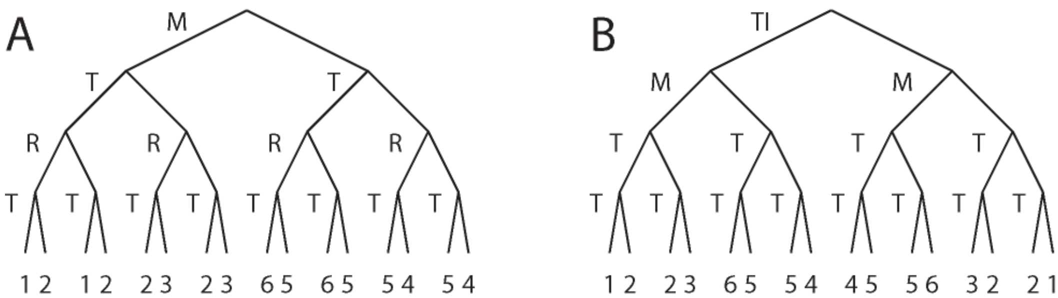


Figure 1.3. Structural trees (adapted from Restle, 1970, 1976). Different operator rules (repeat [R], transpose [T], mirror image [M], temporal inversion [TI]) relate the numbers in the sequence at ascending levels of hierarchical organisation. (A) The two halves of the pattern are related by a single M rule (equivalent to an inverse transformation). (B) The two halves of the pattern are related by a single TI rule (equivalent to a retrograde transformation).

The sequences in Figure 1.3 each have four levels of rule structure, and there is a symmetrical arrangement of rules at these levels. Using Restle's coding system, the sequence in panel A can be recursively rewritten as a simple code incorporating the operation rules repeat (R), transpose (T) and mirror image (M): M(T(R(T(1)))). The sequence in panel B can be recursively rewritten as a simple code incorporating the operation rules transpose (T), mirror image (M) and temporal inversion (TI): TI(M(T(T(1)))). These example patterns are ”ideal” hierarchical patterns because they can be represented by a symmetrical arrangement of rules at all hierarchical levels, and the second half of the pattern can be generated from the first half by a single rule. According to the theory, when they are permitted, hierarchical organisations of patterns are favoured (as opposed to non-hierarchal linear organisations; see Jones & Zamostny, 1975) because they can be represented by a shorter code, and are therefore more economical.

Support for the hierarchical model has come from a number of studies (Fountain & Rowan, 1995; Kundey & Rowan, 2014; Restle & Brown, 1970; Restle, 1970). For example, Restle and colleagues (Restle & Brown, 1970; Restle, 1970) conducted experiments which presented participants with a row of six lights that turned on and off in repetitive sequence. The task was to predict which light would come on next. Analysis of the anticipation error profiles showed that peaks in error rate were associated with transition points between subsequences, and the greatest error rates occurred at higher-level transition points. These results provided evidence for participants’ sensitivity to hierarchical structure, and suggested that the highest-level structure was particularly important. More recently, Fountain and Rowan (1995) have reported that humans (and rats) encode and use multilevel hierarchical structure representations in learning patterned sequences. The task was to reproduce pattern sequences of two, three or four levels of hierarchical structure. Stimuli were presented to human participants on a computer screen and consisted of sequences of objects that appeared at different positions on a circular array. More errors were made for complex patterns (four levels of structure) compared with simpler patterns (two levels of structure). When violations of the hierarchical structure were inserted into sequences, more errors were committed, and these errors were consistent with the rule describing the overall pattern. In other words, participants were sensitive to the hierarchical structure that described the organisation of the majority of pattern elements. This organisation dominated participants' representations and, as a result, inconsistent pattern elements conformed to the hierarchical organisation.

Within the framework of coding theory, subsequent developments were made that highlighted three issues that had not been properly addressed by Restle's hierarchical model (for an early discussion see Jones, 1981). The first concerned the representation of structural relationships at different levels of generalisation. In Restle's coding system, pattern rules operated on interval structure, i.e. the rules specified the direction and specific size of the differences, or "distance", between events on an interval scale. Hence, a transpose rule could be described as being +1 or -1, +2 or -2, and so on. However, Jones (1976, 1981) has demonstrated that the difference between events in a sequence can be described not only at the interval level, but also at the ordinal and nominal levels.[6] An ordinal relation describes the direction of a difference between two events on a scale, without specifying the specific size of the difference, whilst a nominal relation simply describes whether two events are the same or different, without

[6] Though not explicitly acknowledged, Jones' use of the terms interval, ordinal and nominal was presumably informed by Stevens' (1946) theory of scales of measurement.

specifying either the direction or size of the difference. The implication of this is that structural relationships in a sequence can be represented at different levels of generalisation – when a perceiver is unable to represent a pattern at the interval level, they may still be able to represent it at the ordinal level. Support for this view can be found in research addressing melodic processing that has demonstrated a melody can be identified by ordinal structure alone when interval structure has been changed (e.g. Dowling & Fujitani, 1971). The importance of ordinal structure in the processing of sequential patterns will be discussed further in Section 1.5.3.

The second concerned the formalisation of pattern structure. Jones (1981) has argued that Restle's original coding system was incomplete and that a more powerful and coherent rule system was required. Jones (1974, 1978) proposed a system that shares some theoretical properties with the transformational approach (TA) developed to formalise pattern structure in static visual patterns (discussed above): it is also informed by a mathematical approach (see Weyl, 1952) and specifies regularities as those aspects of a sequence that remain invariant under certain transformations. In Jones' system, two groups of symmetry rules were developed that were able to describe all possible relationships in a pattern. The first was called the "Four" group and included the rules identity, complement, transpose and reflection. The complement and reflection rules are broadly equivalent to Restle's mirror image and temporal inversion rules. The second was called the Next group and contained adding and subtracting rules. According to Jones, these group rules permit a representation of pattern structure in terms of different types of relations (nominal, ordinal, interval) and offer a flexible schema for rewriting a series of concatenated rules as a hierarchical structure.

Finally, the third issue concerns temporal relations. Any event in a sequential pattern is defined not only by 'what' it is (e.g. its value on an alphabet scale) but also by the time of occurrence. Restle's coding theory did not account for the fact that events in a sequential pattern can be separated by different inter-onset intervals, may be of different durations and may include periods of silence. In other words, it didn't account for rhythm. Other researchers have therefore developed coding theories that specifically address pattern structure arising from temporal relations (e.g Jones, 1976b; Povel & Essens, 1985; Povel, 1984). The processing of temporal relations and rhythm is discussed further in Section 1.5.3.2.

1.5.3 Auditory pitch patterns

Another area of psychological research that has studied the processing of sequential pattern structure is music cognition. Music is highly patterned and contains a large amount of structural redundancy. For example, once the rhythm of a song is established it will often remain unchanged; the verse of a pop song might involve a 4-bar chord sequence that is repeated over and over again, and choruses will be repeated note for note many times within a song. Even the most complex compositions will usually incorporate reoccurring melodic phrases or themes. Melody, which by its most basic definition is a pattern of sequentially occurring pitches, is arguably the most ubiquitous form of musical structure (Schmuckler, 2009, p.93). As noted in Section 1.3, different types of melodic transformation have been systematically used in music composition, and their processing has received some attention in the psychological research. Before reviewing this work, it is necessary to review some important aspects of melodic structure. A melody consists of structural relationships on two dimensions: pitch and time. The following sections will give a brief overview of how pitch and temporal relations may contribute independently to melody and also how they might be integrated, followed by a review of experiments that have examined the recognition of inverse and retrograde melodic transformations.

1.5.3.1 Pitch relations

The perception of melody necessarily begins with the perception of pitch, which itself can be said to involve some form of pattern recognition (Goldstein, 1973; Shamma & Klein, 2000; Terhardt, 1974).[7] There is a long history of theoretical and empirical research concerned with the perception of pitch (for reviews see Cheveigné, 2004; Yost, 2009), and though there is no definitive model at present, existing models generally agree that pitch is "the perceptual correlate of the periodicity, or repetition rate, of an acoustic waveform" (Oxenham, 2012, p.13335). A periodic waveform can be decomposed into a series of frequency components that are harmonically related through whole number integers to a fundamental frequency (f_0) (e.g. von Helmholtz, 1859/1954). Thus, the pitch of a tone is typically expressed as the frequency of its f_0. Another definition of pitch might be "that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high" (American National Standards Institute: www.ansi.org) – tones are typically placed higher on the scale as their f_0 increases.

[7] Recognition is the identification of something that has been previously learned and is stored in memory. Pattern recognition models of pitch perception propose that the pitch of a sound is determined by comparing the incoming signal with internalized harmonic templates.

Models of pitch perception implicate the physical properties of a tone (such as its f_0) in determining its perceived pitch. This concept of pitch can be more specifically referred to as *absolute pitch*. However, the ability to identify the absolute pitch of a tone, without the use of a reference pitch, is a rare skill (Ross, Gore, & Marks, 2005). In a melodic context, the physical properties of tones are actually less important than their relational properties – in other words, absolute pitch is less important than *relative pitch* (Attneave & Olson, 1971). The essential role of structure in melody has been known for a long time, and was observed by the Austrian philosopher Christian von Ehrenfels (1890/1988, 1937) who viewed melody as a 'Gestalt' percept (structured whole) that is more than just the sum of its individual tones – though the same tones rearranged in time will give rise to a new melody, the melodic Gestalt will remain the same even when all the pitches (order preserved) are transposed to a different pitch register. Interestingly, Ehrenfels compared the perception of melody with that of visual form, pointing out that they both share the property of being transposable without losing their identity – just as visual shapes preserve their identity when translated to different regions of the visual field, melodies preserve their identities when transposed to different pitch registers. The distinction between absolute and relative information is clearly an important one – it is evident that the identity of a pattern is preserved in relational properties. In other words, the essence of a pattern may be found in its *structure* that may be abstracted from the physical properties of a stimulus.

There are a number of structural aspects that arise from relative pitch properties of melody. Of particular importance are pitch interval and pitch contour.[8] Pitch is logarithmically related to frequency, and as a result, pitch intervals are perceptually equivalent (invariant) when the f_0s of their constituent tones stand in the same ratio. The octave is the most recognisable interval, and describes two tones whose f_0s stand in the ratio 1:2. Tones separated by an octave are perceived as being very similar in pitch – a phenomenon known as *octave equivalence* – and this is reflected in the fact that in most musical systems of the world they are considered equivalent and given the same name (Burns & Ward, 2013). This has led psychologists to propose that pitch should be analysed on at least two dimensions: *pitch height*, which describes the overall pitch level; and *pitch chroma*, which describes the position of the pitch within the octave (e.g. Shepard, 1964).

Although interval structure clearly plays a crucial role in melodic processing, ordinal structure is arguably more fundamental (Fujioka, Trainor, Ross, Kakigi, & Pantev, 2004). The configuration of ordinal pitch relations between all of the tones in a melody is called *contour*. Contour plays an important role in both the identification of a melody and its perceived complexity. Empirical studies investigating melodic memory have demonstrated that melody can be identified by contour alone, that is, when interval information has been changed. This was demonstrated by early experiments which found that familiar melodies could be recognised when all of their pitch intervals had been expanded or reduced, by multiplying or dividing all intervals by an equal number, preserving their proportional sizes with relation to each other (Werner, 1925). Another

[8] In Western music, tonality has been the primary determinant of pitch structure in melody throughout recent centuries (Cross, 1985). However, the present research is interested in aspects of melodic structure that are generalizable to non-musical contexts.

experiment involving familiar melodies revealed they can be identified even when all intervals are equalised and set to one semitone (White, 1960).

Subsequent experiments have focussed on memory for novel melodies. Much of this work has adopted the short-term recognition paradigm, in which participants are presented with two melodies (a standard followed by a target), and must discriminate between targets that preserve the structure of the standard and those to which a structural change has been made (Dowling & Bartlett, 1981; Dowling & Fujitani, 1971; Dowling, 1978). In these experiments, changes can be made to local pitch relations (those that occur between two adjacent tones) that change its interval whilst concurrently either preserving or altering its contour. When targets are repeated without transposition, so that structurally identical targets also preserve the absolute properties of tones, both contour-preserving and contour-altering interval changes are easily identified (Dowling & Fujitani, 1971). However, with transposition, targets with contour-preserving interval changes are much more likely to be accepted as being the same as the standard, compared to targets with contour-altering interval changes (Dowling & Bartlett, 1981; Dowling & Fujitani, 1971; Dowling, 1978). These findings suggest that when recognition must rely on relational cues alone (and cannot be based on absolute information), contour structure is more useful than interval structure in identifying a melody. This may be because contour can be processed more easily than interval structure – whereas a representation of interval structure must include both the direction and size of pitch relations, a representation of contour structure describes only the direction of pitch relations. Therefore, a representation of a pattern's contour might be more economical than a representation of its interval structure as it requires fewer demands on memory resources to process. Behavioural results such as those presented above have also led to it being suggested that pitch interval and contour are processed independently (Dowling, 1978). This view has some support in the neuropsychological literature which has identified hemispheric differences associated with the processing of interval and contour information, suggesting that they may be mediated by distinct neural substrates (Liégeois-Chauvel, Peretz, Babaï, Laguitton, & Chauvel, 1998; Peretz, 1990; Schuppert, Münte, Wieringa, & Altenmüller, 2000) The neural correlates of melodic processing will be discussed further in Section 1.7.

Boltz, Marshburn, & Jones (1985) have demonstrated that performance in a melody recognition task can be more dependent on complexity measured by contour reversals than on interval rule structure (see Restle & Brown, 1970; Restle, 1970, discussed above). Participants were presented with melodies which were either hierarchical or not (hierarchical melodies could be represented by a symmetrical rule structure – one half of the pattern was related to the other by a single rule – producing a shorter code than non-hierarchical melodies), and had between 0 and 3 contour reversals (a contour reversal is a point at which the direction of pitch trajectory changes – complexity increases with the number of reversals). The task was to identify which of two comparison melodies was identical to a preceding melody. It was found that rule structure had relatively little effect on performance, though accuracy was reduced significantly as the number of contour reversals increased. This finding was supported by a subsequent study that found contour complexity was also a better predictor of performance in a melody reproduction task than coding complexity (Boltz & Jones, 1986).

It should be noted that the ordinal pitch relations can be organised hierarchically, and this is reflected in models of contour similarity that have placed different emphasis on local and global relations. For example, Quinn's (1999) combinatorial model proposes that mental representations of contour structure include local relations between adjacent and non-adjacent tones. Quinn's model has had some success in predicting participants' subjective judgements of structural similarity (Quinn, 1999). More specifically, though non-adjacent contour relations also contribute, relations between adjacent tones were most significant in determining participants' judgements. Support has also been found for a model of contour similarity that focuses on a more global definition: Schmuckler's (1999) model characterises the rises and falls within the contour through the use of time series analyses, and highlights two important aspects of global contour: 'oscillation', and 'repeated or cyclical patterns'. 'Oscillation' describes the general pattern of upwards and downwards pitch movements, and is quantified by measuring the total and mean pitch distances between reversals in direction of the contour. 'Repeated or cyclical patterns' simply describes the repetition of a particular pattern in a contour, such as a repeated arch-shaped melody, and is quantified using Fourier analysis (both amplitude and phase are thus proposed to be appropriate measures of contour structure). It should be noted that the author acknowledges some difficulties with this approach. In particular, Fourier analysis requires the sampling of an infinitely periodic signal, but melodic contours are short and discrete – as a result the analysis is particularly influenced by information at the beginning and end of the melodic pattern. To test the model's validity, Schmuckler (1999) carried out an experiment in which listeners rated 12-note melodies for subjective complexity on a scale of 1 (not very complex) to 9 (very complex). These ratings were then used to derive a measure of perceived contour similarity, which was compared with the measures derived from Schmuckler's model. Correlational and regression analyses revealed that there was some predictive power for the measures based on the degree of oscillation and cyclical information.

Taken together, the above evidence suggests that melodic contour can be described using models based on both local (Quinn, 1999) and global (Schmuckler, 1999) parameters. Schmuckler (2009) has suggested this might reflect the fact that there is not one single mechanism responsible for the abstraction of contour at both the local and global levels. Rather, perception of a global contour may be a result of a combination of low-level processing, subject to higher-level integration.

1.5.3.2 Temporal relations

A melody unfolds over time. Therefore, any melodic pitch relation is inherently temporal, and it follows that any consideration of melodic structure must include both pitch and temporal relations. In the music psychology literature, these have traditionally been treated separately (Justus & Bharucha, 2002; Krumhansl, 2000). However, their independence has been questioned by other researchers, who hold that the perception, attention and memory for pitch relations is inherently rhythmical, and as such, listeners treat melody and rhythm as a unified dimension (Ellis & Jones, 2009; Jones, Johnston, & Puente, 2006; Jones, 1974, 1987).

Much progress in understanding the human capacities for processing temporal information was made by experimental investigations carried out by

Fraisse and colleagues (for a review see Fraisse, 1984), who have defined time as a psychological notion that refers to two different concepts: succession, which corresponds to the fact that two or more events can be perceived as different and organised sequentially, and duration, which applies to the specific time interval between successive events. In summarising Fraisse's work, Krumhansl (2000) concluded that, though it has been demonstrated that humans can quite accurately estimate absolute time and detect small differences in duration, the most impressive abilities are found in the perception and production of rhythms. Therefore, the patterns of durations, rather than absolute durations, seem to be psychologically primary.

Rhythm is commonly defined as the temporal patterning of event durations in an auditory sequence. A fundamental phenomenon for the perception of rhythm is the perception of a periodic pulse (Fraisse, 1982, 1984). The perception of a periodic pulse, or beat period, can provide an important anchor in the time domain, as it offers a temporal frame or perspective for the perception of other time relations (multiplications or subdivisions of a specified beat period). Many different theoretical approaches have been adopted in order to explain how the temporal relations among elements in a serial pattern are represented (Buhusi & Meck, 2005; Buonomano & Karmarkar, 2002; Creelman, 1962; Essens, 1986; Lerdahl & Jackendoff, 1983; Povel & Essens, 1985; Treisman, 1963). Those approaches rooted in the tradition of absolute time have not been particularly successful in explaining the perception of rhythmically structured sequences (e.g.

centralised clock models: Creelman, 1962; Treisman, 1963).[9] Just as relative pitch is more important in melody recognition than absolute pitch, the relative durations between event onsets is more important for rhythm recognition than their absolute durations. Accordingly, the temporal units of rhythmic patterns are more appropriately expressed as ratios of the beat period. For example, a temporal pattern of durations 500ms, 500ms, 1000ms can be expressed as the following ratios (assuming the perceived beat period would be equal to the smallest time unit in the sequence): 1:1, 1:1, 1:2. As recognition is based on this relative information, simple rhythms that share the same ratio durations but different absolute time durations are recognised as being equivalent. Models specifically designed to explain perception (and production) of rhythmic patterns on the assumption that durations are represented as ratios of a beat period have taken different approaches to the encoding of successive time intervals in a sequence (e.g. Lerdahl & Jackendoff, 1983; Povel & Essens, 1985; Povel, 1981). Perhaps the most influential (and arguably less music-centric, and hence more generally applicable to non-musical temporal structure) has been Povel's clock model (Povel & Essens, 1985; Povel, 1984). Briefly, this conceives of rhythmic organisation as involving a temporal grid of intervals generated by an internal clock or pacemaker. The conceived internal clock is hierarchical, in that it entrains to the pulse of a temporal pattern, but it does not simply match the smallest

[9] It has been argued that temporal models based on network dynamics may be better suited to handling the higher-order structured organisation of complex stimuli such as music (e.g. Buhusi & Meck, 2005; Buonomano & Karmarkar, 2002). From this perspective, distinct networks may be involved in temporal processing depending on the task and modality being used. Accordingly, cortical networks are proposed to be inherently able to process temporal information because information about recent input history is captured by time-dependent changes in the state of the network.

duration of a sequence. Rather, multiple clocks are conceived that generate pulses at different units of time, encompassing different periods that fit within the pattern, and different locations (phases).

Although melodic and rhythmic structure have generally been treated separately in the psychological literature, Jones and colleagues have argued that time and rhythm function psychologically as integral parts of the unfolding melody, and that melodic change is not independent of the time intervals and temporal context in which a pitch change transpires (Boltz et al., 1985; Boltz, 1991; Ellis & Jones, 2009; Jones et al., 2006; Jones, Summerell, & Marshburn, 1987; Jones, 1987). They have developed a dynamic attending theory (Jones & Boltz, 1989; Jones, 1976b; Large & Jones, 1999), according to which temporal patterns in real-world events can synchronise attending via the mechanism of entrainment. Entrainment is the physical process whereby internal attending periodicities become attuned to salient recurrent stimulus time spans. The resulting attentional synchronies are possible at multiple time scales, and are facilitated when time spans at different time scales are hierarchically nested. In auditory patterns such time spans are marked by accents that arise from salient serial changes in pitch, called melodic accents, and/or timing, called temporal accents. When melodic and temporal accents are present in a single pattern, they contribute to the emergence of a common higher order time structure, called *joint accent structure* (JAS; Boltz & Jones, 1986; Jones, 1987).

Following work by Thomassen (1982), Jones has identified two types of melodic accent: a pitch-contour accent, which depends on a temporal ordering of pitches and results from a local change in direction of pitch trajectory, with accentuation on the inflection point; and a pitch-leap accent, which falls on the second of two tones that form a pitch interval that is larger than the preceding pitch intervals in a series. The temporal succession of melodic accents (or melodic accent pattern) can contribute to a higher order temporal pattern even in the absence of temporal accents, when a melody is isochronous. A temporal accent refers to any salient local serial change in time relationships, which can arise from a change within a serial pattern of inter-onset intervals, or when a tone has a longer duration than neighbouring tones in a sequence (Ellis & Jones, 2009).

Support for this approach to studying dynamic pattern structure comes from a number of studies (Boltz & Jones, 1986; Ellis & Jones, 2009; Jones et al., 2006; Jones, Moynihan, MacKenzie, & Puente, 2002; Jones et al., 1987). Perhaps most relevant to the present research is a recognition study by Jones et al. (1987) which demonstrated the importance of the relative timing of pitch trajectories associated with contour (i.e. pitch-contour accents), and named this temporal contour structure the *dynamic shape*. In a learning phase participants were familiarised with a number of melodies that differed with respect to pitch relationships and rhythm. Then, in a test phase, participants were presented with targets that included the previously learned melodies, interspersed with decoys that either shared the same contour (but different interval structure) or did not. In addition, all targets were presented either in the original rhythm or a new rhythm. On hearing each target melody, participants indicated whether the melody was old or new. The results from Experiment 1 indicated that decoys with the same contour were more likely to be confused with old melodies than decoys with different contours, and that decoys were most confusing when they shared the same contour and rhythm (i.e. dynamic shape). Jones and colleagues interpreted these findings in terms of their dynamic attending theory in which remembering is assumed to involve recapitulation of the original rhythmical activities involved in attending to melodies.

1.5.3.3 The perception of melodic transformations

The perception of global structural regularities has been well researched in the visual domain (e.g. see symmetry perception, discussed in Section 1.5.1). Whilst classic work in serial pattern learning research has demonstrated that observers are able to use global structural regularities to represent sequential patterns, relatively little work has been carried out to investigate how efficiently the different types of regularity described by inverse and retrograde transformations are perceived. Some researchers in music cognition have taken an interest in these transformations in a melodic context, due to their use in musical composition (see Section 1.3). The present section will review some of these studies.

Perhaps the first experiment to test listeners' ability to recognise melody under retrograde transformation was carried out by White (1960). Participants were presented with well-known melodies (either 6 notes or 24 notes in length) that had undergone different types of distortion and were played in both forward and reverse order. The task was to indicate the identity of the melody from a multiple-choice list. Though participants were able to identify melodies that had undergone a retrograde transformation at better than chance level, they did not find the task easy, and performance was no better than when pitch information had been removed entirely from the melodies, so that responses were made on rhythm information alone.

Using the short-term recognition paradigm, Dowling (1972) conducted an investigation into the recognition of transformations of novel 5-note, isochronous melodies. Participants were presented with a standard melody that consisted of randomly generated notes from the chromatic scale.[10] Target melodies were either related (under transformation) or unrelated (a different randomly generated melody) to the standard, and that started on a tone randomly chosen from the range of an octave above or below the first tone of the standard. Three groups of participants were required to discriminate between related targets that had undergone either an inverse, retrograde, or retrograde inverse transformation and unrelated targets. Nested within the related condition was a structural change variable (unchanged, contour-preserving interval change). The rate of presentation was either 5 tones per second or 2 tones per second and, for exact transformations presented at the fast rate, tasks were performed either with or without a preceding analogous visual task. Finally, participants were instructed to either accept exact transformations only as being the same as the standard, or to accept any contour-preserving target.

Although participants recognised all three types of transformation better than chance level (>50% correct), there was some variation in performance between conditions, with inversions being the easiest transformation to identify (mean accuracy 70%), followed by retrogrades (mean accuracy 64%) and then retrograde inversions (mean accuracy 55%). Inversions were recognised better than chance under all conditions, and retrograde under all except when required to identify exact transformations presented at the faster rate. Retrograde inversions

---

[10] The chromatic scale is used as a basis for all musical scales in the Western music tradition. It divides the octave into twelve equally tempered intervals (or semitones).

were recognised better than chance level only when participants were required to identify contour-preserving transformations at the slower rate, and exact transformations at the faster rate when preceded by the visual task. In general, the melodies were harder to recognise when presented at the faster rather than the slower rate. It was also harder to recognise exact transformations of melodies compared with contour-preserving transformations. Exact interval size information seemed to be lost with transformation – suggesting that contour information rather than discrete intervallic information is important for recognition (in agreement with transposition recognition experiments; Dowling & Bartlett, 1981; Dowling & Fujitani, 1971; Dowling, 1978).

More recently, Cupchik et al. (2001) tested participants' ability to discriminate melodic targets that were either exact transformations of a standard (inverse, retrograde), from targets that were transformations which included a contour-preserving interval alteration. Melodies were isochronous and either 3 or 7 notes in length. Participants were more accurate in identifying exact retrograde transformations compared to exact inverse transformations. However, there were some methodological issues with this experiment. Unlike Dowling (1972), Cupchik et al. did not transpose target melodies. As a result, exact retrograde transformations preserved the absolute pitches of notes that were presented in the standard melody. The availability of this information may have facilitated performance, whilst discrimination of targets under inverse transformation could only have relied on structural information. Secondly, exact and inexact target melodies both preserved the contour of the standard – thus Cupchik et al.'s results may reflect participants' ability to identify changes to interval structure rather than contour, which has been previously demonstrated to be particularly important to melodic processing.

There is some limited evidence to suggest that retrograde and inverse transformations of melodies may be recognised without explicit instruction. With explicit instruction participants are informed of the type of transformation they are required to process. Without explicit instruction, participants may process the transformations implicitly, i.e. without conscious awareness of the relationship between transformed melodies. Balch (1981) carried out a study in which participants rated different melody transformations for ‘good continuation’. ‘Good continuation’ relates to the Gestalt principle which states that component parts or objects tend to be grouped together when one follows the established configuration of the other (Koffka, 1935). Participants were presented with isochronous sequences 10 or 16 tones in length. The first half of the sequence was randomly generated from a diatonic scale, with no one pitch repeated. The second half of the sequence was either one of the three types of transformation (inverse, retrograde, retrograde inverse), or unrelated. Participants simply rated how well the second melody continued on from the first. Retrograde was rated highest for good continuation, followed by inversion, and they were both rated significantly higher than unrelated. Although ratings for retrograde inversions were also higher than ratings for unrelated, this difference failed to reach statistical significance. These results suggest that, without explicit instruction, retrograde inverse transformations were too hard to detect and therefore did not contribute to ratings of good continuation. The finding that retrograde transformations were rated as ‘better’ continuations than inverse transformations contrasts with Dowling’s (1972) finding that inverse transformations were easier to recognise. However, whilst retrograde targets in Dowlings's study had been transposed to a different pitch register, in Balch's study they constituted an untransposed reversal of tones. Therefore, it is possible that good continuation ratings of retrograde transformations were informed not only by structural information (relative pitch) but also by physical properties of the tones (absolute pitch).

The results of Dowling (1972) and Balch (1981) suggest that listeners are able to recognise transformations of short melodies, and that the structural relationships between transformations may be recognised, to some extent, implicitly. Krumhansl et al. (1987) have shown that listeners find it harder to accurately identify transformations of longer and more rhythmically complex melodies. In an attempt to examine whether listeners perceive any similarity between a tone row and its transformations in the serialist style of music, the authors carried out an initial experiment in which musically trained listeners familiarised themselves with two different isochronous 12-note tone rows (row 1 and row 2) in an initial training phase. In the test phase, the same tone rows and transformed versions (inversion, retrograde, retrograde inversion) were presented in random order. Participants were informed of the different types of transformation that had been applied to the test stimuli, and their task was to indicate whether the tone row sounded more like row 1 or 2. Analysis of the results, excluding the identical form, showed overall accuracy was better than chance level (approximately 70-80%) though there was no significant effect of transformation type.

The experiment was repeated with tone rows involving more rhythmic and melodic variety. Once again, overall performance was above chance level, with no significant differences between transformation types. However, closer inspection of individual participants' data revealed that nearly half did not reliably perform above chance level. It should be noted that in both experiments, there was a great amount of individual difference – not all participants were always successful at accurately labelling the transformation. Performance was generally correlated with musical training. There was also an interaction between two groups of participants who had previously been judged to respond differently to the sequences in a probe-tone task. One group (which had received more training on average) appeared to find retrograde inversion judgments hardest, whilst the opposite was true of the other group. This suggests that the ability to process of global transformations of melody can be developed with musical training.

A more recent experiment investigating serialist/12-tone music found that participants with no specialised training could not implicitly learn the structural relationships between transformed melodies (Dienes & Longuet-Higgins, 2004). Stimuli used were tone rows, the second half of which were transformations of the first half (transposed, inverse, retrograde, retrograde inverse). Four groups of participants took part and were exposed to just one type of transformation. In the learning phase, participants rated tone rows for pleasantness. They were then told that the stimuli they had heard obeyed some set of rules, and that half of the stimuli that they were about to hear (in the test phase) would obey the same rules and half would not. The task was to classify the test stimuli. Accuracy scores were at chance for all stimuli types, suggesting that participants did not implicitly learn the structural relationships.

In contrast, when a similar follow-up experiment was carried out, participants with routine exposure to and interest in serialist music were found to classify stimuli with 56.3% accuracy, which was found to be significantly better than chance. In addition, analysis of confidence ratings that accompanied responses suggested that participants believed they were guessing, regardless of how accurate they were, lending support to the assumption that perception of the transforms was implicit rather than explicit.

In summary, the results of the experiments reported above demonstrate that inverse and retrograde transformations are easier to process in isolation compared to a combination (retrograde inverse transformation). Following a transformational approach (discussed in Sections 1.5.1 and 1.5.2), which formalises regularities in terms of a set of transformations, it may be assumed that retrograde inverse transformations are harder to process because they involve two transformations as opposed to only one. The transformational approach would also predict that inverse and retrograde transformations are equally recognisable, as they each require a single transformation. But this is not always the case. Separate experiments have demonstrated that either of the two may be easier to recognise than the other. It appears that inverse transformations are easier to process when recognition is based on structural information only (possible reasons for this are discussed in Chapter 2). On the other hand, when absolute pitch information is available, it appears that retrograde transformations are easier to process. However, further investigation is required to confirm this observation.

1.6 Pitch, time and space

As it has already been noted, inverse and retrograde transformations are not exclusive to melody, but describe regularities that are found in sequential pattern structure generally (see Sections 1.3 and 1.5.2). Given the distinction that has been made between sensory information and the structural information that must be abstracted from it in order to perceive pattern (see Section 1.4), it may be hypothesised that cognitive representations of structural information are supramodal, i.e. they transcend specific sensory modalities, and may be subject to supramodal cognitive processes. In connection with this hypothesis, a number of theoretical and empirical studies point towards the possibility that melodic structure is represented spatially (Eitan & Granot, 2006; Eitan & Timmers, 2010; McLachlan, Greco, Toner, & Wilson, 2010; Rusconi, Kwan, Giordano, Umiltà, & Butterworth, 2006). The implication of this is that structural information that has been abstracted from melodic stimuli in the auditory domain or from visuo-spatial stimuli in the visual domain, is encoded as a supramodal spatial representation that is neither auditory nor visual. In turn, this would imply that the processing of melodic and visuo-spatial patterns involves a supramodal mechanism (or mechanisms) sensitive to spatially represented structural information.

This concept of supramodal structural space is distinct from concepts of space associated with spaciousness perception and localisation. It can be related to the theoretical perspective offered by Kubovy (1988). In discussing theoretical analogies between visual space and auditory space, Kubovy has asserted that auditory space (in terms of sound localisation and spaciousness perception) is not a direct analogue of visual space. Instead, he suggests that visual and auditory analogies should be sought in Gestalt phenomena whereby perceptual organisation in vision occurs in space and time, and perceptual organisation in audition occurs in pitch and time. In this sense, although the auditory characteristic of pitch cannot be said to be inherently spatial, it dominates spatial location in determining the perceptual organisation of sound.

Studies addressing the spatial representation of structural information do not sit within the mainstream of psychological research, and there has not been a sustained and focussed attempt to explore this topic thoroughly. However, for the purposes of the present thesis, relevant research will be reviewed by placing them into two broadly distinct categories: those that are concerned with the spatial representations of simple stimuli of different pitch and timing, and those that are concerned with the spatial representation of more complex melodic stimuli.

1.6.1 Spatial representations of time and pitch

In psychological research, there is growing evidence for the spatial representation of inherently non-spatial stimuli on various dimensions, including time and auditory pitch (Dehaene, Bossini, & Giraux, 1993; Evans & Treisman, 2010; Fischer, 2003; Gevers, Reynvoet, & Fias, 2003; Ishihara, Keller, Rossetti, & Prinz, 2008; Lakens, Semin, & Garrido, 2011; Lidji, Kolinsky, Lochy, & Morais, 2007; Rusconi et al., 2006; Shaki & Fischer, 2008). Some researchers have suggested that the spatial representations associated with time and pitch (and other 'non-spatial' dimensions such as numbers and letters of the alphabet) reflect a general spatial representation of ordinal information (Gevers et al., 2003; Ginsburg & Gevers, 2015). Similarly, others have argued for the existence of shared magnitude representation (Cohen Kadosh & Henik, 2006; Cohen Kadosh, Lammertyn, & Izard, 2008).

The timing and pitch of events has been associated with a spatial representation that corresponds to the horizontal and vertical axes of visual space. A common paradigm that has been used to investigate spatial representations of time and pitch is the stimulus-response compatibility (SRC) paradigm. An SRC effect occurs in speeded choice reaction tasks, when there is dimensional overlap between stimuli and responses, influencing the speed and accuracy of performance. Dimensional overlap between stimuli and responses can occur in their spatial location, so that stimuli appearing higher in vertical space are responded to faster with an upper response key, irrespective of whether the keys lie on a frontal or on a transverse plane (Cho & Proctor, 2003; Vu, Proctor, & Pick, 2000). Using this paradigm, Ishihara et al. (2008) demonstrated a spatial representation of time that corresponds to a horizontal axis. Participants were presented with a probe stimulus following periodic auditory clicks, and pressed one of two keys depending on whether the timing of a given probe was earlier or later than expected, based on the previous clicks. When response buttons were aligned horizontally, left-sided responses to early onsets were faster than those to late onsets, and right-sided responses to late onsets were faster than those to early onsets. When response buttons were aligned vertically, no SRC effect was observed.

The spatial representation of time has also been demonstrated in studies that have used alternatives to the SRC paradigm. Lakens et al. (2011) conducted two experiments to examine how spatial representation of auditory and visual time may converge, in terms of how time is structured in visual space and how it is structured in auditory space. Their first experiment required participants to order words with temporal meaning on a horizontal line, without any specific instruction. As predicted, past-related words were placed significantly further to the left than future-related words. Further analysis showed that placements were not dichotomously organised, but reflected a continuous representation. In a second experiment, an implicit task was employed whereby the same words (and non-temporal control words) were presented binaurally over headphones and participants indicated whether the left or right signal was loudest. Future-related words presented at the same loudness were judged to be louder in the right ear. There was a cross-modal overlap between the results observed in both experiments: the likelihood of words being judged as being louder in the right ear was correlated with the placement of words on the horizontal line.

Whereas time is associated with a spatial representation on the horizontal axis, pitch height is strongly associated with a spatial representation on the vertical axis. When instructed to represent tones with different pitches in two-dimensional visual space, pitch height is systematically mapped onto vertical height, with higher pitches being represented as being higher in visual space, and lower pitches as being lower (Mudd, 1963). This tendency to map pitch height onto a vertical scale may even dominate our ability to locate sounds being emitted from different positions in vertical space (Pratt, 1930). Associations between pitch height and vertical height may occur automatically and at the perceptual level. Experiments investigating cross-modal correspondences, in which participants are required to classify vertically high or low visual objects that are accompanied by tones that are either high or low in pitch (or vice versa), have found that response times are faster when visual and auditory stimuli are congruent (i.e. the participant is faster to classify an object that is high on the screen when accompanied by a tone that is high in pitch, compared with a low pitch; Bernstein & Edelstein, 1971; Evans & Treisman, 2010). This correspondence persists when the task is indirectly related to pitch and vertical height, when participants are required to classify the timbre of the sound or the surface pattern of the visual object (Evans & Treisman, 2010).

A correspondence between pitch height and vertical height has also been revealed by experiments employing the SRC paradigm. When classifying the pitch height of two tones, an SRC effect has been observed between pitch height and response location – participants are quicker to classify pitch height when ‘high pitch’ responses are allocated to an upper response key and ‘low pitch’ responses are allocated to a lower response key, compared with when ‘high pitch’ responses are allocated to a lower response key and ‘low pitch’ responses are allocated to a higher response key (Lidji, Kolinsky, Lochy, & Morais, 2007; Rusconi, Kwan, Giordano, Umiltà, & Butterworth, 2006). This SRC effect persisted when the task was indirect and participants were instructed to classify sounds by timbre.

While the association of pitch height and vertical height appears to be most dominant, it should be noted that pitch height has also been associated with space along the horizontal axis, with low pitch to the left and high pitch to the right. Mudd (1963) found that tones with different pitches were not only systematically displaced on the vertical axis, but also (though to a lesser degree) on the horizontal axis. In addition, SRC effects have been found when classifications of high and low pitches are made using response buttons aligned on the horizontal axis, however this effect was not found to persist when the task was indirect (Lidji et al., 2007; Rusconi et al., 2006).

1.6.2 Spatial representations of melodic structure

The empirical research reported above has demonstrated seemingly clear and systematic cognitive associations between the inherently non-spatial dimensions of time and pitch, and visual spatial dimensions. These studies used simple stimuli, but that does not rule out the possibility that the observed associations are linked with structural processing – though the experiments typically collected responses to isolated events, the task always required participants to make a comparison with some reference stimulus. For instance, responses to tones of different pitch in the SRC paradigm required participants to judge whether the target tone was the higher or the lower of two possibilities (Lidji et al., 2007; Rusconi et al., 2006). Therefore, responses were based on representations of relative pitch as well as absolute pitch.

Other studies have investigated the possibility that shared processes are involved in the perception of more complex melodic and visuo-spatial patterns. O'Leary and Rhodes (1984) demonstrated that when auditory and visual sequences are presented simultaneously, perceptual organisation in one sensory modality may influence organisation in the other. Auditory streaming refers to the phenomenon where sequences of tones with different pitches may be grouped into a single stream or segregated into different streams, according to bottom-up cues such as temporal proximity and similarity in f_0 (Bregman, 1990). More proximal and more similar sounds tend to be perceived as a single stream, while less proximal and less similar sounds tend to be perceived as multiple streams. Between these two points, stream perception may be ambiguous and may depend on top-down influences such as attention, expectations and prior exposure. In O'Leary and Rhodes' experiment, participants were presented with pitch sequences and analogous visual sequences which, instead of tones with different pitches, consisted of small rectangles presented at different heights on a computer screen. First, auditory sequences were presented independently. The pitch distance between tones and the tempo of presentation was manipulated to discover the parameters at which participants perceived the sequences either as a single integrated stream, two segregated streams, or an ambiguous stream. Independently presented visual sequences were also manipulated to achieve the same effect by manipulating the vertical distance between rectangles and the tempo of presentation. When an ambiguous auditory sequence was paired with a segregated visual sequence, it was perceived unambiguously as two separate streams. When the same ambiguous auditory sequence was paired with an integrated visual sequence it was unambiguously perceived as a single stream. This cross-modal influence was found to occur in both directions.

Other experiments have investigated the possibility of a general mechanism being involved in the abstraction of global contour. Contour is not an exclusive attribute of melody, and it has been shown that contour information can be accurately abstracted from sequential patterns in other auditory dimensions such as loudness and timbre (McDermott, Lehr, & Oxenham, 2008). Using a short-term recognition paradigm, Balch and Muscatelli (1986) presented participants with short, isochronous melodies and analogous visual sequences. Visual sequences consisted of discrete objects presented at different vertical heights, and unfolded from left to right on a computer screen. Standard and target patterns were combined in four presentation conditions: auditory-auditory (A-A), auditory-visual (A-V), visual-auditory (V-A), visual-visual (V-V); and the rate of presentation was also manipulated. While there was no difference in accuracy of recognition across conditions at the fastest rate of presentation, recognition performance for V-V conditions improved significantly relative to A-A conditions as the presentation rate slowed. The authors concluded that auditory contour is harder to abstract than visual contour. Interestingly, cross-modal conditions also improved significantly compared to A-A, suggesting that whilst visual contour is abstracted more readily than auditory contour, contour structures perceived in either modality are easily relatable to each other. In a more recent experiment, Prince, Schmuckler and Thompson (2009) examined participants' similarity ratings of melodies (long and short) and analogous visual line drawings that were either matching or had been slightly altered. The ratings for matched contours exceeded those for mismatched contours and provided further support for cross-modal sensitivity to contour. They further demonstrated that similarity ratings could be predicted by a model of global contour shape, originally developed for application to melodic contour (discussed in Section 1.5.3.1; Schmuckler, 1999).

A very limited number of studies have investigated possible spatial processing associated with inverse and retrograde transformations of melodic structure. As reported earlier, it has been shown that performance in a melodic transformation recognition task was facilitated when melodies were accompanied by an analogous visual representation (Dowling, 1972). More recently, a study by McLachlan, Greco, Toner and Wilson (2010) demonstrated that performance in a melodic transformation recognition task is facilitated more by strictly spatial compared to symbolic visual representations. The authors pointed out that traditional music notation combines both spatial and symbolic representations of pitch, and argued that if pitch is encoded spatially then a notation which relied on the spatial properties of pitch should facilitate cognitive performance in recognition tasks involving melodic transformation. For the purposes of their research, a graphic notation was developed that depicts pitch in a more spatially coherent way. Participants were required to identify melodic transformations with or without the aid of either traditional music notation or the newly developed graphic notation.

In each experimental trial, participants were presented with melody pairs. Melodies were either presented as auditory only, or they were accompanied by visual notation (classic or graphic notation). The second melody was either different to the first melody, or it had undergone a transformation (transposed, inverse, retrograde). Participants were given two tasks, in some sessions they were required to indicate whether the second melody was the same or different to the first (discrimination task). In other sessions they were required to identify the specific transformation that had been applied to the second melody (manipulation task). Unsurprisingly, the discrimination of melodic contour was easier than recognising its transformation. In addition, the simultaneous presentation of music notation facilitated performance in all tasks, with the spatially congruent graphic notation being especially effective. The authors concluded that this supports the proposition that pitch is spatially encoded in auditory short-term memory.

Finally, it was reported above that Cupchik et al. (2001) found that retrograde transformations of melody are recognised more accurately than inverse transformations. The authors carried out additional tests and discovered that accurate judgments of retrograde transformations predicted accuracy at performing visuo-spatial mental rotation tasks (e.g. Shepard & Metzler, 1971), which they argued provides evidence for shared processes in spatial rotation and melodic transformation. In order to explain why performance in the spatial task was predicted by accuracy of retrograde and not inverse judgements, they suggested that listeners are not easily able to reverse the temporal order of notes in a melody, and for this reason would have relied on a more holistic strategy that preserves the global shape of a melody when identifying retrograde targets. In contrast, they suggested that inverse comparisons encourage a sequential local comparison of the degree to which successive notes are mirror images of the standard. However, there are some issues with this study. Firstly, it was correlational, providing arguably weak results. Secondly, as noted earlier, target melodies were not transposed, and thus recognition of retrograde targets may have been facilitated by absolute pitch information.

1.7 Neural correlates and hierarchical processing in the brain

Further evidence in support of the hypothesis that transformations of structure abstracted from the auditory and visual domains may involve shared mechanisms may be found in neuropsychological research. Briefly, information processing in the brain is thought to obey two main principles: hierarchical processing and functional specialisation. Hierarchical processing refers to the way in which perception is achieved via gradual steps in which information is first represented in a localised simple form and through a sequence of processes is transformed into more abstract, holistic, and even multimodal representations (DeYoe & Van Essen, 1988). Though hierarchical processing has historically been modelled as a bottom-up process with local analysis preceding global analysis (Hubel & Wiesel, 1968), other models have demonstrated how global analysis can occur in parallel and therefore guide local analysis (Bullier, 2001). Functional specialisation refers to the existence of specialised anatomical areas that process information about different aspects of the perceptual scene. The visual and auditory systems may consist of parallel hierarchical sequences, or processing streams, that are specialised for a particular functional task (Barrett &

Hall, 2006; Creem & Proffitt, 2001; Goodale & Milner, 1992; Rauschecker & Tian, 2000; Rauschecker, 2013; Ungerleider & Haxby, 1994; Ungerleider & Mishkin, 1982; Zatorre, Bouffard, Ahad, & Belin, 2002).

Thus, the processing of pitch information occurs in the auditory cortex (Johnsrude, Penhune, & Zatorre, 2000; Tramo, Shah, & Braida, 2002; Warrier & Zatorre, 2004; Zatorre, 1988), which is anatomically distinct from the visual cortex where simple features of visual objects are analysed (Bartels & Zeki, 2000; Tootell & Hadjikhani, 2000; Tootell et al., 1995; Watson et al., 1993). However, as analysis moves from local processing to global processing, activity spreads out, away from these sensory specific areas, possibly along functionally distinct information pathways (Barrett & Hall, 2006; Creem & Proffitt, 2001; Goodale & Milner, 1992; McLachlan & Wilson, 2010; Rauschecker & Tian, 2000; Rauschecker, 2013; Ungerleider & Mishkin, 1982; Zatorre et al., 2002). Eventually, these pathways converge at higher order multimodal brain areas in the temporal, parietal and frontal cortices. It is at these points that the hypothesised supramodal mechanisms responsible for the inverse and retrograde transformation of structure might be found. In support of this assumption, recent brain-imaging studies have identified areas of the parietal cortex (specifically, the intraparietal sulcus [IPS]) that are associated with the processing of melodic transformations (Foster, Halpern, & Zatorre, 2013; Foster & Zatorre, 2010a, 2010b; Zatorre, Halpern, & Bouffard, 2010). The following section will give an overview of hierarchical processing and functional specialisation in the visual and auditory systems, before discussing the parietal cortex as a potential site for supramodal mechanisms in more detail.

In the visual system, object perception begins when visual sensory information is encoded in the retina and is sent via the lateral geniculate nucleus (LGN) to the primary visual cortex (V1), also called the striate cortex, situated laterally in the occipital lobe. Retinotopic maps are preserved in early visual areas, including V1, V2, V3, V4/V8 and V3a (Wandell, 1999). The processing of visual features such as colour, motion and depth have been associated with additional processing in later areas such as middle temporal (MT or V5) and medial superior temporal (MST) areas (Bartels & Zeki, 2000; Tootell & Hadjikhani, 2000; Tootell et al., 1995; Watson et al., 1993).

It is commonly held that processing proceeds along functionally distinct pathways that project from V1 to higher order brain areas (see Figure 1.4). The two main pathways identified are the dorsal and ventral streams, which project to the inferotemporal cortex and the posterior parietal cortex, respectively (Creem & Proffitt, 2001). The ventral stream is often called the “what” stream and corresponds to object or form recognition (Goodale & Milner, 1992; Ungerleider & Mishkin, 1982). The dorsal stream is often called the “where” stream and corresponds to spatial processing (Ungerleider & Mishkin, 1982). There is still some debate regarding the function of the streams, and some have argued that object and spatial processing take place in both streams (Creem & Proffitt, 2001; Goodale & Milner, 1992). In agreement with this perspective, visual object recognition has been associated with a large constellation of areas, in both dorsal and ventral pathways, that lie anterior and lateral to early retinotopic areas (for a review see Grill-Spector & Malach, 2004). Grill-Spector and Malach (2004) have discussed the organisation of these areas in terms of three subdivisions: lateral occipital cortex (LOC), ventral occipitotemporal (VOT) regions and dorsal loci.

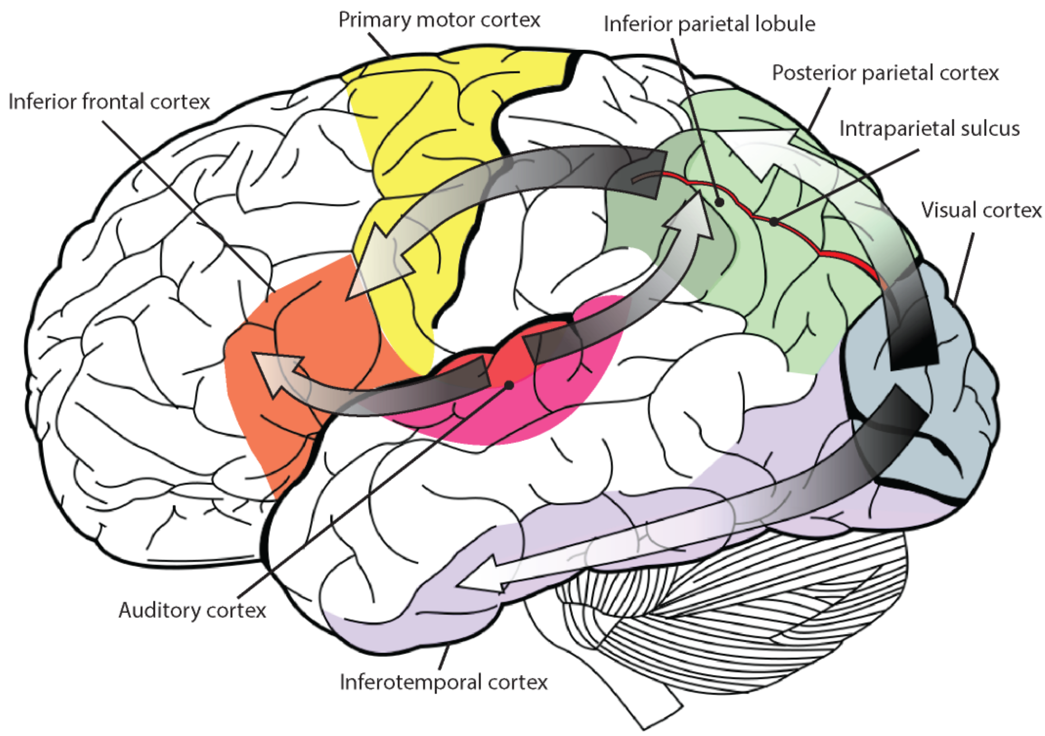


Figure 1.4. Simplified model of processing streams (adapted from Creem & Proffitt, 2001; Rauschecker, 2013). NB intraparietal sulcus marks boundary between superior parietal lobule associated with higher order spatial processing of visual information and inferior parietal lobule associated with higher order processing of melody.

In comparison with the visual system, the processing in the auditory system has received much less attention, and is comparatively poorly understood. The processing of auditory pitch begins when sound is encoded in the peripheral auditory system. Within the cochlea, hair cells embedded along the basilar membrane transduce physical vibrations into electrical activity in nerve fibres, which is transmitted to the brain. Information ascends the auditory pathway via intermediary stages (brainstem, midbrain and thalamus) before being relayed to the auditory cortex, located in the temporal lobe. According to a recent review by Moerel, De Martino and Formisano (2014), the human auditory cortex is generally situated bilaterally on the supratemporal plane and comprises two thirds of the superior temporal gyrus (STG). It can be divided into three anatomical regions. In anterior to posterior direction, it includes planum polare (PP), the transverse temporal gyrus or Heschl's gyrus (HG), and planum temporale (PT). Formally, the auditory cortex was divided into primary (A1) and secondary (A2) projection areas and further association areas. The modern divisions of the auditory cortex are the core area (which includes A1), a surrounding belt region, and a parabelt region which is situated posterior-laterally to the lateral portion of the belt region (de la Mothe, Blumell, Kajikawa, & Hackett, 2006; Hackett, Preuss, & Kaas, 2001; Hackett, Stepniewska, & Kaas, 1998a, 1998b). The tonotopic organisation of frequency, encoded in the cochlea, is maintained up to A1 (Kaas, Hackett, & Tramo, 1999) and perhaps beyond (Talavage et al., 2004). It has been shown that neurons in the core areas of monkeys and humans respond strongly to narrowband sounds (such as pure tones), and neurons in the belt areas respond better to complex sounds (both periodic and non-periodic) (Petkov, Kayser, Augath, & Logothetis, 2006; Wessinger et al., 2001).

Reflecting developments in visual research, functionally specialised pathways have also been proposed to explain processing in the auditory system (see Figure 1.4; Barrett & Hall, 2006; McLachlan & Wilson, 2010; Rauschecker & Tian, 2000; Rauschecker, 2013; Zatorre et al., 2002). Rauschecker and Tian (2000) initially proposed a ventral stream that specialises in object recognition (and was hence called the "what" stream) that projects from the auditory cortex towards the posterior parietal cortex (PP) and prefrontal cortex (PFC), and a dorsal stream that specialises in localising the spatial position of auditory objects (and was hence called the "where" stream) that projects from the auditory cortex towards anterior temporal areas before converging with the dorsal stream in PFC.

However, a recent reassessment by Rauschecker (2013) remodels the ventral stream as an anteroventral stream that projects to the inferior frontal cortex (IFC) via the anterior superior temporal regions, and the dorsal stream as a posterodorsal stream that projects to the primary motor cortex (PMC) via the inferior parietal lobule (IPL). Whilst the role of the anteroventral stream in auditory object recognition is presently widely accepted, the role of the posterodorsal stream continues to be debated (Rauschecker, 2013).

The importance of the human auditory cortex for the representation of pitch has been confirmed by a number of studies in which patients with lesions in the auditory cortex have demonstrated impaired pitch discrimination abilities (Johnsrude et al., 2000; Tramo et al., 2002; Warrier & Zatorre, 2004; Zatorre, 1988). In addition, the right hemisphere appears to play a particularly important role, as patients with lesions that encroach on HG in the right hemisphere have been associated with more pronounced impairments in pitch direction judgments of both pure and complex tones (Johnsrude et al., 2000; Warrier & Zatorre, 2004; Zatorre, 1988). An fMRI study by Patterson, Uppenkamp, Johnsrude and Griffiths (2002) has identified higher order areas that may be associated with melodic processing. Sounds were presented to participants that matched spectrally, but either produced no pitch, a fixed pitch, or a melody. It was found that all conditions activated areas in HG and PT. However, sounds which communicated a perceptible pitch increased activation only in the lateral half of HG. Furthermore, once the sound began to vary in pitch, producing a melody, activation was recorded in regions beyond HG and PT. Varying pitch was correlated with activity in STG and PP. These results support a hierarchical model of pitch processing – and demonstrate that as the complexity of the stimulus increases, activity moves anterolaterally away from A1.

Neurological studies of relative pitch processing have found further hemispheric differences, which suggest that local interval and global contour information may be mediated by distinct neural substrates (Liégeois-Chauvel et al., 1998; Peretz, 1990; Schuppert et al., 2000). For instance, Liégeois-Chauvel et al. (1998) found that damage to the right temporal cortex impaired the processing of both contour and interval information in the discrimination of melodies, whilst damage to the left temporal cortex impaired only the processing of interval information. As a result of findings such as these, a cortical hierarchy model has been proposed in which global processing of contour occurs in the right hemisphere and acts as a ‘‘framework’’ on which the local detail is subsequently hung when interval information is analysed in the left hemisphere (e.g. Peretz, 1990). Partial support for this model has come from an fMRI study that examined the brain areas associated with the processing of pitch changes that either preserved or violated the contour of melodic sequences (Stewart, Overath, Warren, Foxton, & Griffiths, 2008). When the two conditions were compared, additional activation was observed in the right PT and posterior superior temporal sulcus (pSTS). However, contrary to previous studies, the processing of contour changes was associated with activity in the left hemisphere, while the processing of interval changes was associated with bilateral activity. Another recent fMRI study has investigated the higher order areas of the brain that may contribute to the processing of pitch contour. Lee, Janata, Frost, Hanke and Granger (2011) found that three areas are associated with discriminating between ascending and descending melodic contour – the right STS, the left inferior parietal lobule (IPL) and the anterior cingulate cortex (ACC).

Taken together, the neurological and brain-imaging data reported above are in agreement with the notion that increasingly abstracted auditory information is associated with processing in functional areas of the brain spreading away from core areas of the auditory cortex and towards higher order multimodal areas. One higher-order region of the brain that is of particular interest is the parietal cortex, and more specifically the intraparietal sulcus (IPS). Recent brain-imaging studies by Zatorre and colleagues (Foster et al., 2013; Foster & Zatorre, 2010a, 2010b; Zatorre et al., 2010) have associated areas of the parietal cortex with the processing of global transformations of melodic structure. In particular, Zatorre et al. (2010) demonstrated that the IPS was activated by the mental reversal of familiar melodies (i.e. retrograde transformation). A subsequent study investigated the areas associated with the discrimination of novel transposed and reversed melodies (Foster et al., 2013). Five-note tonal melodies were presented that were either exact replications or had had a contour preserving interval change made. The recognition of both transposed and reversed melodies was associated with activity in the bilateral IPS. As reported by the authors, the pattern of activation observed in the IPS may be compared with activation in the same area when other mental transformation tasks are performed, such as mental rotation of visuo-spatial patterns, quantitative calculation and visually-guided manual tasks (Alivisatos & Petrides, 1997; Frey, Vinton, Norlund, & Grafton, 2005; Harris et al., 2000; Ischebeck et al., 2006; Kong et al., 2005; Zacks & Michelon, 2005; Zacks, 2008). Based on Zatorre and colleagues' findings it is proposed that the IPS and its surrounding cortical areas may serve as the neural correlate for the hypothesised supramodal mechanisms responsible for the processing of melodic transformations.

1.8 Summary and orientation of the thesis

Pattern perception is a fundamental function of the human perceptual system, and underlies many psychological processes. A perceived pattern is a structured whole, or Gestalt, which is made up of interrelated parts and components. The perceptual system appears to be driven towards finding patterns in the environment to achieve the simplest possible interpretation. One advantage of this is that it minimises the expenditure of energy: patterns signal redundancy, redundant information can be predicted, and in turn it does not need to be processed to the same extent as previously encoded information. Pattern structure can be formalised as a set of regularities that describe invariance under transformation at different hierarchical levels (local and global). Global regularities are of particular use to the perceptual system because they signal the greatest proportion of redundant information, thus saving the greatest amount of energy.

The perception of global regularities has been well researched in the visual domain, with a focus on static patterns. Less well researched is the perception of global regularities in sequential patterns. Classic research in this area has demonstrated that observers are able to use global regularities to represent complex sequences (Restle & Brown, 1970; Restle, 1970), however, many developments in our understanding of sequential pattern perception come from research that has addressed melody. For example, research in melodic perception has highlighted the importance of relative versus absolute information – a melodic pattern preserves its identity when it is transposed, changing absolute properties but preserving relative properties (Dowling & Bartlett, 1981; Dowling & Fujitani, 1971; Dowling, 1978). This work has highlighted the important roles of interval structure and contour – transposed melodies that share the same contour are more likely to be mistakenly identified as being the same, suggesting that contour plays an important role in the perception of global regularities.

A limited number of studies have investigated the perception of two special types of structural transformation that are found in music: inverse and retrograde (and their combination retrograde inverse). These have confirmed the importance of contour in perception, but provide conflicting results with regard to the perceptual salience of different types of transformation: one study has demonstrated that inverse transformations are processed more easily (Dowling, 1972), another has demonstrated that retrograde transformations are processed more easily (Cupchik et al., 2001). This may have something to do with the way in which retrograde transformations have been applied to stimuli: in the former, transformation did not preserve the absolute pitches of tones, but in the latter it did. Thus, it is possible that the availability of absolute information facilitates the perception of retrograde transformations.

Pattern structure is an abstract property that transcends any particular stimulus (Pomerantz & Lockhead, 1991). Some researchers have proposed that the perception of global pattern structure (in particular, contour) involves supramodal processing mechanisms (Balch & Muscatelli, 1986; Cupchik et al., 2001; O'Leary & Rhodes, 1984; Prince et al., 2009; Williamson, Cocchini, & Stewart, 2011). In relation to this, recent work has demonstrated that auditory pitch patterns may be represented spatially (Evans & Treisman, 2010; Lidji et al., 2007; Rusconi et al., 2006), suggesting the existence of partly shared representations or processes for auditory pitch patterns and visuo-spatial patterns. In support of this notion, brain-imaging data has identified shared neural correlates in higher order cortical areas that are activated by the mental transformation of melodic and visuo-spatial stimuli (Foster et al., 2013; Foster & Zatorre, 2010a; Zatorre et al., 2010).

Taken together, this informs a general hypothesis that was explored in the present thesis: the perception and cognition of auditory pitch patterns involves a general mechanism (or mechanisms) that is responsible for the processing of supramodal structural information. In order to explore this hypothesis, the experiments reported in this thesis examined the processing of inverse and retrograde structural transformations of auditory pitch and visuo-spatial patterns. Inverse and retrograde transformations provide a potentially useful tool for investigating the hypothesised supramodal processes because they transcend specific sensory modalities and engage structural processing mechanisms. The aim was to investigate how these transformations are handled when presented in different unimodal and cross-modal conditions.

Chapter 2: Theoretical framework and general methods

2.1 Introduction

Although pattern perception in the auditory and visual sensory domains is typically studied independently, the evidence presented in Chapter 1 points towards partly shared representations and processing of auditory pitch patterns and visuo-spatial patterns. The primary objective of the present research was to explore this possibility within a newly developed theoretical framework, detailed in the present chapter. The theoretical framework was examined using two types of isomorphic transformation of global pattern structure: inverse and retrograde.[11] These transformations provide a potentially informative means of studying supramodal processes, but to date have received relatively little attention in psychological research. Thus, the experiments reported in the subsequent chapters represent an attempt to initiate a thorough and systematic examination of the processing of inverse and retrograde transformations in unimodal and cross-modal tasks, with a view to analysing response patterns in the context of the conceived theoretical framework.

In order to achieve the research objective, a theoretical framework was developed in which structural information, abstracted from auditory and visual sensory information, determines cognitive representations of patterns in a flexible supramodal pattern space. As a starting point, the framework is intended to apply specifically to isochronous (equal time interval) sequences of tones varying in pitch height and isochronous sequences of visual objects varying in spatial position. From this framework, a hypothesis was generated that was tested over a series of experiments (Experiments 2, 3, 4, 5, 6, 7 and 8), reported in Chapters 4,

[11] These transformations are isomorphic because patterns under these types of transformation have corresponding or similar form and relations.

5 and 6. In the remaining sections of the present chapter, the initial theoretical framework will be outlined in more detail. The hypothesis (and its predictions) will also be outlined, before giving a detailed description of the general methodology employed.

2.2 Supramodal pattern space: a theoretical framework

In this section, a framework is provided for a flexible supramodal pattern space, in which pattern representations can be constructed from one or a combination of two qualitatively distinct types of supramodal dimension. Any supramodal mechanisms involved in the processing of inverse and retrograde transformations are envisaged to operate on these supramodal representations. Before introducing the supramodal pattern space, a useful distinction is made between *non-structural* and *structural* information. Non-structural information refers to the perceived properties of a pattern that correspond to physical attributes of its components, without taking into account the way in which the pattern components are related. Structural information, on the other hand, refers to the perceived properties of a pattern that correspond to the abstracted relationships between its components. To illustrate, in auditory pitch patterns non-structural information would refer to the pitch height of tones that correspond to their fundamental frequency, whereas structural information would refer to the relationships between tones, such as their relative pitch height and the pitch interval between them. In visuo-spatial patterns non-structural information would refer to the spatial positions of objects that correspond to their location in space, whereas structural information would refer to the relationship between objects at different locations, such as their orientation with respect to one another and the relative distances between objects. The information specified within the conceived supramodal pattern space derives from perceived structural relationships, which is abstracted from non-structural, sensory-specific information.

It is here proposed that patterns can be represented in a supramodal space, constructed from one or a combination of qualitatively distinct types of supramodal dimension. Two dimensions are described here: a bidirectional *scalar* dimension, and a unidirectional *temporal* dimension. Table 2.1 gives some examples of different supramodal pattern spaces that can be constructed from scalar and temporal dimensions, and the types of patterns that would be represented in them.

Table 2.1

Examples of different supramodal pattern spaces

Pattern space	Examples
1T	Punctiform rhythmic patterns (e.g. metronome)
1S	Punctiform symbolic patterns (e.g. Morse code)
1ST	Pitch patterns, moving 1S patterns
2S	Visual images (e.g. photograph)
2ST	Moving 2S images (e.g. film)
3S	Static 3S patterns (e.g. sculpture)
3ST	Experienced reality

Note. S = scalar dimension, T = temporal dimension

Simple auditory pitch patterns can be represented in a supramodal pattern space constructed from one scalar dimension and one temporal dimension, which will hereinafter be referred to as a ‘one-and-a-half-dimensional’ (1½-D) pattern space – see Figure 2.1 for an illustration of a pattern represented in a 1½-D space.[12] On the y-axis, the scalar dimension represents the relative pitch height of tones. Ascending pitch height corresponds to ascending values on the scalar dimension. Thus, an auditory event that has a scalar value ‘tag’ of A is lower in pitch height than an event that has a scalar value ‘tag’ of B, and so on. On the x-axis, the temporal dimension represents the relative timing of auditory events. The order of events unfolds from left to right. Thus, an auditory event with the temporal value ‘tag’ of 1 comes before an auditory event with the temporal value ‘tag’ of 2, and so on. Scalar and temporal dimensions are discrete, reflecting the nature of the sequential patterns under investigation. An incremental step along an axis represents a single interval unit. Thus, a distance of two intervals separates the values A and C on the scalar dimension, and a distance of four intervals separates the events 1 and 5 on the temporal dimension. Filled squares on the grid signify pattern events – in this case, tones of the pitch sequence.

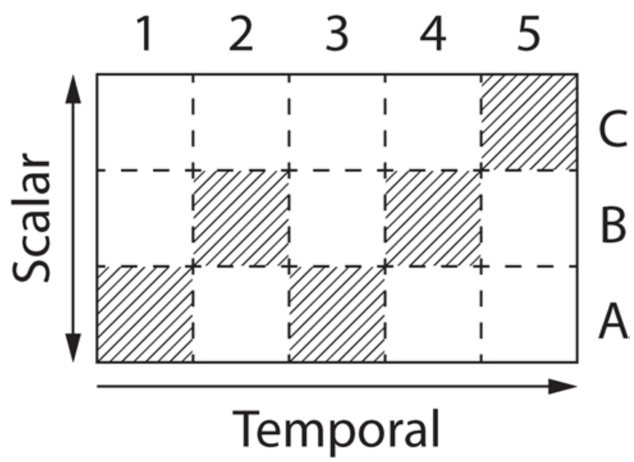


Figure 2.1. Pattern represented in a 1½-D supramodal pattern space.

[12] The temporal dimension is treated as a half dimension here because of its directionality.

Pattern representations in a 1½-D supramodal pattern space can also derive from structural information abstracted from visual stimuli. The scalar dimension represents the relative pitch of auditory tones, but it can represent relative information from any sensory dimension on which pattern components can move freely in either direction. Therefore, it can also represent the relative spatial position of visual objects on dimensions such as vertical height. The temporal dimension represents the relative timing of auditory tones, but it can also represent the relative timing of visual events. So, a visual stimulus that consists of visual objects presented sequentially at different vertical heights would theoretically also correspond to a representation in a 1½-D supramodal pattern space.

The motivation for representing relative pitch height and relative vertical height on the same supramodal spatial dimension originates from two sources, which have been discussed in Chapter 1. Firstly, theoretical and empirical data from serial pattern learning research demonstrate that the same structural rules govern the cognitive organisation and representation of patterns abstracted from visuo-spatial sequences and auditory pitch sequences (Deutsch & Feroe, 1981; Restle, 1970). Secondly, a growing body of empirical research demonstrates auditory pitch is represented spatially (Evans & Treisman, 2010; Lidji, Kolinsky, Lochy, & Morais, 2007; Rusconi, Kwan, Giordano, Umiltà, & Butterworth, 2006). This spatial association is particularly strong on the vertical axis, with higher pitches being represented at higher vertical positions and lower pitches at lower vertical positions.

Representing the timing of auditory and visual events on the same dimension is superficially easier to justify, as the dimension of time transcends sensory modalities. However, on the basis of empirical and theoretical data, this justification is not so straightforward. Although the traditional view has been that time perception involves a centralised clock (Creelman, 1962; Treisman, 1963), empirical data have not always supported this view. For example, it has been demonstrated that interval discrimination is generally better in the auditory than in the visual modality, and interval discrimination between modalities is significantly worse than within modalities (Grondin & Rousseau, 1991; Rousseau, Poirier, & Lemyre, 1983). Furthermore, many studies have shown that entrainment to rhythmic sounds (e.g. tapping) is done with far greater accuracy than to temporally equivalent visual stimuli (for a recent review see Iversen, Patel, Nicodemus, & Emmorey, 2012). Findings such as these are more consistent with an alternative view, which models timing as being processed by distributed networks which depend on the task and modality being used (Buhusi & Meck, 2005; Buonomano & Karmarkar, 2002). Nevertheless, in the context of the conceived supramodal pattern space, it is assumed that temporal relations abstracted from the timing of auditory and visual events are in some sense equivalent.

Another issue concerns the motivation for representing relative timing on a spatial dimension, and why this spatial dimension is unidirectional and not bidirectional (like the scalar dimension). A spatial-temporal association has been reliably observed where the timing of events corresponds to spatial representations on a horizontal axis (Ishihara, Keller, Rossetti, & Prinz, 2008; Lakens, Semin, & Garrido, 2011), with events that are earlier in time represented to the left and events that are later in time to the right. However, there is a fundamental difference between dimensions such as auditory pitch or visual height, and the dimension of time. The dimensions of auditory pitch height and visual vertical height are inherently bidirectional – as a sequence unfolds, events may move up or down these dimensions. Time, on the other hand, is inherently unidirectional – sequences encountered in the environment move forwards in time, but they never move backwards in time. Our mental constructs are largely shaped by our interaction with the environment (Lakoff & Johnson, 1980), suggesting that mentally moving backwards in time should be harder than mentally moving forwards in time. Whilst mental time travel is possible, in the sense that we can think of past and future events (Miles, Nind, & Macrae, 2010), this is not the same as reversing the flow of time and imagining events unfolding in reversed temporal order.

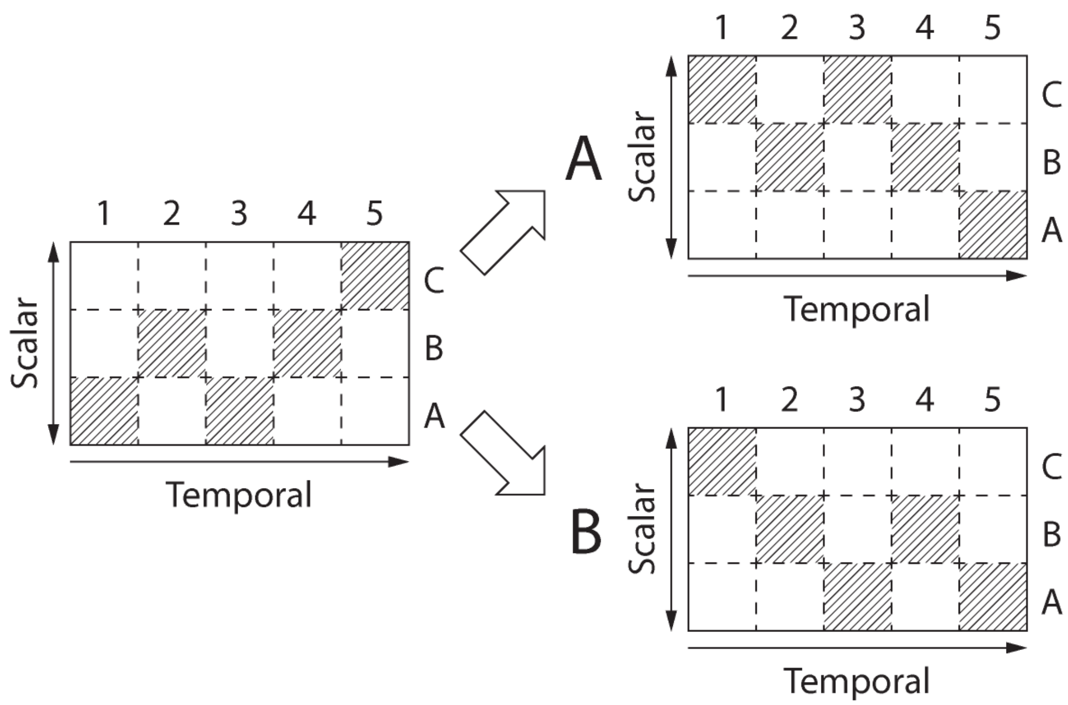


Figure 2.2. Transformations of patterns in a 1½-D supramodal pattern space. (A) Inverse transformation. (B) Retrograde transformation.

Figure 2.2 displays inverse and retrograde transformations in the 1½-D supramodal pattern space. Within the framework outlined here, inverse and retrograde transformations may be formalised according to the pattern dimension that needs to be transformed. Before discussing this in more detail, it is necessary to describe the structure of patterns. Three types of structure have been identified that contribute to representations of sequential patterns: interval, ordinal and nominal (Jones, 1976). Whilst all of these structural levels must play a role in the processing of structural transformations, the present research focuses on the processing of ordinal structural information, as it has been shown in melody recognition experiments that contour is more salient than interval information when discriminating between same and different transformations (Dowling, 1972). The example pattern can be described by the ordinal relations between events on the scalar dimension (1 is lower than 2/2 is higher than 1; 2 is higher than 3/3 is lower than 2; and so on) and on the temporal dimension (1 is before 2/2 is after 1; 2 is before 3/3 is after 2; and so on). When considering the scalar and temporal dimensions in combination, the ordinal relation between successive events can be described by a local scalar-temporal transition (the transition between 1 and 2 is upwards; the transition between 2 and 3 is downwards; and so on). The configuration of all of these locally described ordinal relations determine the global contour description of a pattern. Table 2.2 displays pattern structure descriptions of the example pattern and its transformations displayed in Figure 2.2, demonstrating how local and global relations are altered with transformation. Note that this description of global contour simply serves the purpose of distinguishing between patterns, and it is acknowledged that there are many different ways of formalising psychological representations of contour (for an overview of different models, see Müllensiefen & Wiggins, 2011).

Table 2.2

Description of local (scalar-temporal relations) and global (contour) pattern structure

	Scalar-temporal relations[a]				Contour[b]
	1 to 2	2 to 3	3 to 4	4 to 5	
Example	Upwards	Downwards	Upwards	Upwards	UDUU
Inverse	Downwards	Upwards	Downwards	Downwards	DUDD
Retrograde	Downwards	Downwards	Upwards	Downwards	DDUD

Note. [a] Numbers correspond to events in a pattern; [b] U = upwards, D = downwards

Returning to the formalisation of structural transformations within the theoretical framework, inverse transformation may be formalised as an inversion (or reversal) of ordinal relations on the y-axis, which is a scalar dimension. Retrograde transformation may be formalised as an inversion (or reversal) of ordinal relations on the x-axis, which is a temporal dimension. In turn, it is envisaged that any mechanism that contributes to the perception of pattern regularities described by these types of structural transformation, will involve a process that inverts ordinal relations on these dimensions. Figure 2.3 illustrates how such a mechanism might operate. The transformation process can be conceptualised as a global process that reverses the polarity of the scalar (for an inverse transformation) or the temporal (for a retrograde transformation) dimension, which reassigns the 'tags' associated with pattern events on those dimensions.

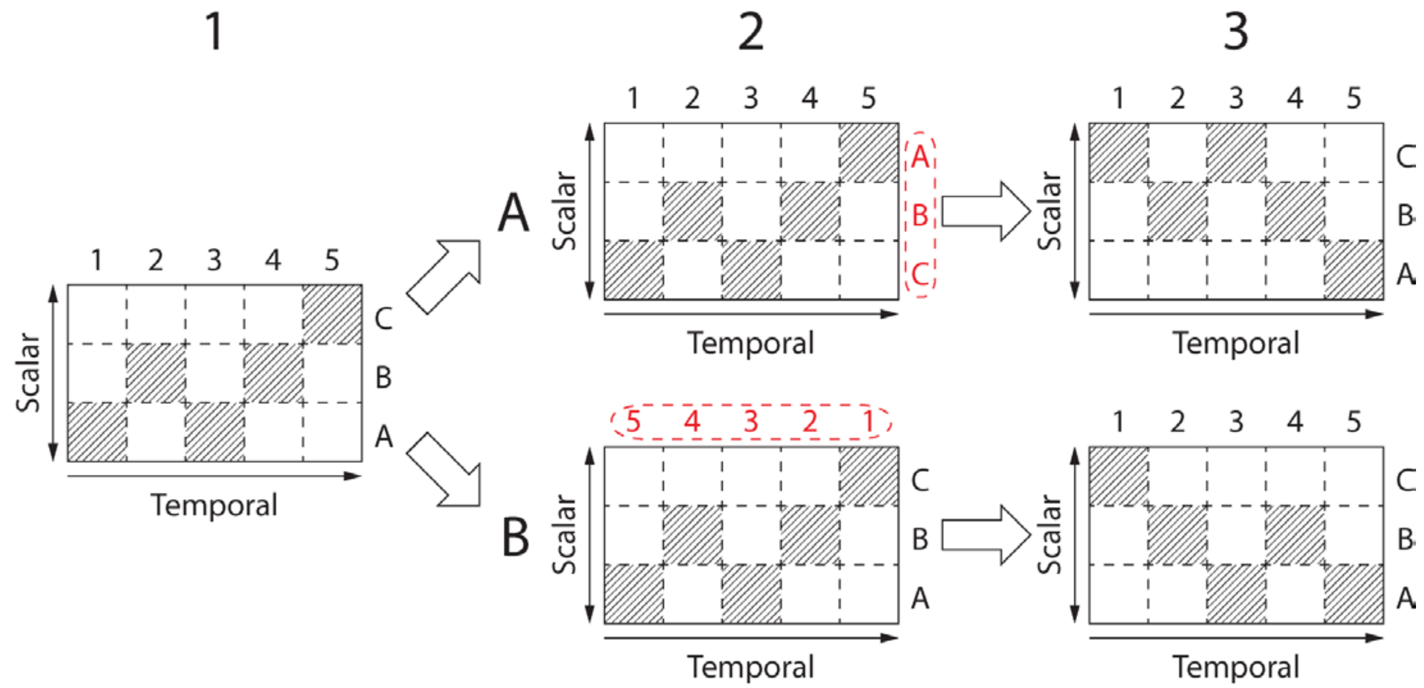


Figure 2.3. Hypothesised transformation process in three stages. 1) An example pattern represented in 1½-D supramodal pattern space. 2) The transformation process for inverse (A) and retrograde (B) transformations. In general, the transformation process involves an inversion (or reversal) of ordinal relations on one dimension, and can be conceptualised as a global process that reverses the 'polarity' of the dimension, resulting in the reassignment of value 'tags' on the relevant dimension (highlighted in red). Inverse transformation involves an inversion of ordinal relations on the scalar dimension. The polarity reversal on the scalar dimension is compatible with its bidirectionality. (B) Retrograde transformation involves an inversion of ordinal relations on the temporal dimension. The polarity reversal on the temporal dimension is incompatible with its unidirectionality. 3) The transformed pattern.

As argued above, representations on the scalar and temporal dimensions are fundamentally different – ordinal relations on the scalar dimension are bidirectional, but ordinal relations on the temporal dimension are unidirectional. It can be said that representations on the temporal dimension are therefore not as spatial as those on the scalar dimension, owing to the inherent directionality of ordinal relations on this dimension. A consequence of this is that the process of inverting ordinal relations on both dimensions is not equally compatible. An inverse transformation, which requires the inversion of ordinal relations on a scalar dimension, is relatively compatible due to the bidirectionality of the dimension (as demonstrated in Figure 2.3A). In contrast, a retrograde transformation, which requires the inversion of ordinal relations on a temporal dimension, is relatively incompatible with its directionality (as demonstrated in Figure 2.3B). Importantly, this suggests that retrograde transformations would be processed less efficiently than inverse transformations.

2.2.1 The 'one-and-a-half-dimensional' (1½-D) hypothesis

Following the supramodal pattern space (SPS) framework outlined above, a 1½-D hypothesis was defined and tested in a series of experiments reported in Chapters 4, 5 and 6. The hypothesis was based on the assumption that auditory and visual stimuli can be represented on equivalent supramodal dimensions, and that the perception of pattern regularities, described by inverse and retrograde transformation, involves the processing of structural information on these supramodal dimensions. As described in the previous section, auditory pitch patterns can be represented in a 1½-D supramodal pattern space, constructed from a scalar and a temporal dimension. The scalar dimension represents relative pitch, and the temporal dimension represents the relative timing of tones. Visuo-spatial patterns can also be represented in a 1½-D supramodal pattern space, when the scalar dimension represents relative vertical height and the temporal dimension represents the relative timing of visual objects. Thus, in order for representations of visual patterns to be equivalent to representations of auditory patterns, a visual stimulus should consist of objects presented sequentially at different positions, on a single spatial dimension. To be most closely analogous, objects should be distributed along the vertical dimension, as this has been most strongly associated with representations of pitch. Visual stimuli were presented this way in an experiment carried out by O'Leary and Rhodes (1984) who examined cross-modal effects on auditory and visual object perception, demonstrating that perceptual organisation in one modality could influence perceptual organisation in the other (this experiment was discussed in more detail in Chapter 1, Section 1.6.2).

At present, very little is known about the perceptual processes underlying inverse and retrograde transformations. To the author's knowledge, only a handful of studies have investigated the relative perceptual salience of both of these transformations (Cupchik, Phillips, & Hill, 2001; Dowling, 1972; Krumhansl, Sandell, & Sergeant, 1987). These were all concerned with melodic transformations in the auditory domain, and provide conflicting data. For a pattern represented in a 1½-D space, an inverse transformation involves an inversion of ordinal relations on the scalar dimension, and a retrograde transformation involves an inversion of ordinal relations on the temporal dimension. One of the assumptions of the theoretical framework is that transformations of ordinal structure on the temporal dimension are incompatible with its inherent directionality, and therefore harder to process. This predicted that pattern relationships under retrograde transformation should be perceptually less salient than pattern relationships under inverse transformation, regardless of the sensory modality in which patterns are presented.

In the auditory domain, there is some evidence to support this hypothesis. As discussed in Chapter 1 (Section 1.5.3.3), Dowling (1972) has demonstrated that inverse transformations of melody are recognised more accurately than retrograde transformations. Although there has been some research that shows that sequential visuo-spatial patterns under inverse and retrograde transformation can be perceived by an observer (Restle & Brown, 1970; Restle, 1970, 1976), to date there have been no studies that specifically address the relative salience of inverse and retrograde transformations, either in unimodal or cross-modal conditions. Previous studies that have made a structural analogy between auditory pitch patterns and visuo-spatial patterns have either not investigated inverse and retrograde transformations (Balch & Muscatelli, 1986; O'Leary & Rhodes, 1984; Prince, Schmuckler, & Thompson, 2009), or have simply demonstrated that structurally analogous visuo-spatial patterns facilitate performance in a melodic transformation task (Dowling, 1972; McLachlan, Greco, Toner, & Wilson, 2010).

A summary of the 1½-D hypothesis can be seen in Table 2.3. In short, it predicted that when patterns are represented in a 1½-D supramodal pattern space, inverse transformations should be perceived more effectively than retrograde transformations, regardless of the sensory modality from which structural information is abstracted.

Table 2.3

Summary of the 1½-D hypothesis

Hypothesis	Assumptions	Prediction
1½-D	• Structural information, abstracted from auditory and visual stimuli, can be represented in a supramodal pattern space, constructed from a bidirectional scalar dimension and a unidirectional temporal dimension. • Inversions of structural information on the temporal dimension are processed less effectively than inversions of structural information on the scalar dimension, due to the former's inherent directionality. • The scalar dimension represents the relative pitch of tones or the relative spatial position of visual objects. The temporal dimension represents the relative timing of auditory or visual events. • Simple auditory pitch patterns are represented in a 1½-D supramodal pattern space, constructed from a scalar and a temporal dimension. • Inverse transformations of patterns represented in 1½-D space require an inversion of ordinal relations on the scalar dimension. Retrograde transformations of patterns represented in 1½-D space require an inversion of ordinal relations on the temporal dimension.	The perception of pattern regularities described by inverse transformation will be more effective than the perception of pattern regularities described by retrograde transformation, regardless of the sensory modality in which patterns are encountered.

2.3 The experimental paradigm

One way to investigate how inverse and retrograde transformations of sequential pattern structure are processed by the perceptual system would be to examine performance in a recognition task. The short-term recognition paradigm has been used extensively in the melodic processing literature (e.g. Bartlett & Dowling, 1980; Dowling & Fujitani, 1971; Dowling, 1971, 1972; Edworthy, 1985; Gosselin, Jolicoeur, & Peretz, 2009), and was adopted for the experiments which comprised the main body of the research (Experiments 2, 3, 4, 5 and 6).[13] The paradigm involves experimental trials in which pairs of stimuli are presented one after the other. The first stimulus (hereinafter referred to as the *standard*) is attended to by the participant and retained. After a short pause, a second stimulus (hereinafter referred to as the *target*) is presented which is either the same as the standard (related) or different from it (unrelated). Once the target has been presented the participant must indicate whether they recognise it to be the same as the standard or different (what may be considered the same is subject to the experimenter's instructions).

The paradigm is particularly appropriate for the proposed research as it allows the presentation of a variety of stimuli in both unimodal and cross-modal conditions. It also permits extensive manipulation of stimulus parameters such as duration and rate of presentation. In addition, other aspects of a trial's design such as durations of the inter-stimulus interval (ISI) and response interval may also be controlled. Alternative techniques such as a recall paradigm would not have been

[13] Different experimental paradigms were used in Experiments 1, 7 and 8. The paradigms used in these experiments (and other methodological details) are described in the relevant chapters of the thesis (Chapters 3 and 6).

appropriate as responses would have been difficult for participants to produce and problematic for the experimenter to analyse. For example, when recalling melodies it may be difficult for a participant with no musical expertise to reproduce what they have heard. A matching task, in which a participant is presented with multiple targets and required to identify which one is related to a standard stimulus, was not deemed appropriate due to the temporal nature of stimuli. It would be difficult to retain the standard pattern in short-term memory over an extended period of time. It would also mean the experiment would become prohibitively long.

Returning to the present thesis, in each experimental trial of experiments adopting the short-term recognition paradigm, related targets shared the same pattern structure as the standard, having undergone either an inverse or a retrograde transformation (see Figure 2.12 for the time course of experimental trials; this will be discussed further in the other sections of the chapter). Taking into account the added difficulty of comparing patterns across modalities, it was anticipated that retrograde inverse transformations would be too difficult for participants to identify. This was confirmed by the findings of a pilot experiment. Unrelated targets were structurally different and were not related to standards under inverse, retrograde, or retrograde inverse transformations. The main dependent variables used in the current experiments were: 1) error rate, operationalized as percentage error (PE), and 2) correct response time (RT).

Both measures are accepted ways of indirectly examining conscious mental processes and are used extensively in the areas of sensation, perception and cognition. PE permits the study of a system that is revealed by its failures when overloaded or otherwise taxed. Thus, the sensory process of interest is investigated by varying the level of some factor (in this case transformation and modality) and examining its effects on the pattern of PE. In contrast, RT permits the study of a system when it is functioning well. The processing of information in the brain is recognised as being highly structured - different pathways through that structure result in different time courses, which are revealed by differences in RTs (Luce, 1986). Thus, from RT data one is able to make inferences about the structures involved under different experimental conditions. Due to the difficulty of the task in the present research, high PE was recorded relative to typical experiments where RT is the principal dependent variable, and therefore PE data were given special treatment (see Section 2.9). In general, processing efficiency is reflected in error rates and response latencies – targets that are processed less efficiently are revealed by larger error rates and longer response latencies; targets that are processed more efficiently are revealed by smaller error rates and shorter response latencies.

2.4 Stimuli

The design of the stimuli began with the production of a pool of supramodal pattern structures. As the present research seeks to better understand the cognition of auditory pitch patterns, any decisions made about stimulus parameters were made from an auditory perspective first before being applied to visual stimuli.

The pattern structures and stimuli described in the following section were used in Experiments 2, 3, 4 and 5. In Experiment 6 the pattern structures and stimuli were slightly modified – these modifications are described in the relevant methods section of Chapter 5.

2.4.1 Generation of the stimulus structure

Pattern structures used in Experiments 2, 3, 4 and 5 were computed from a 5x3 matrix (e.g. Figure 2.4). Values on the x-axis represented the timing of events and values on the y-axis represented different pitches or spatial positions. In the previous literature adopting a short-term recognition-memory paradigm to investigate the cognition of pitch sequences, stimuli of various lengths have been employed, usually in the range of 3 notes (e.g. Cupchik et al., 2001) to 7 notes (e.g. Cuddy & Lyons, 1981) in length. For the purposes of the present research it was necessary to employ patterns that could be held in short-term memory (STM) efficiently so that participants would be able to perform the relevant mental transformation. Due to the limited capacity of STM (Berz, 1995; McConnell & Quinn, 2004; Mukari, Umat, & Othman, 2010) it would not be possible to perform the task with patterns that are too long or complex. At the same time, patterns that are very short or simple allow only a limited number of unique patterns to be generated. Patterns of length five were thus chosen as it has been shown that pitch sequences of this length are easily encoded and retained in STM (Dowling & Fujitani, 1971), can be mentally transformed (Cupchik et al., 2001; Dowling, 1972), and can be compared with analogous visual sequences (Balch & Muscatelli, 1986; McLachlan et al., 2010). This length also meant that a sizeable number of unique patterns could be formulated.

A total 243 patterns were initially computed. In order to select only the patterns that were suitable for the purposes of the research, those that did not meet the following criteria were discarded. First, all patterns were required to include one of the three possible values (A, B and C) at least once. As a result, all patterns were ternary. Though this constraint has not been set in previous research, it ensured a high degree of informational consistency (all patterns were the same length and included three different pitches/vertical heights). Second, all patterns had to be *transformationally distinct* (i.e. unique and not equivalent to a transformation of another pattern). For example, the pattern CBABA was the same as the retrograde transformation of ABABC, meaning that it was not transformationally distinct and was therefore discarded (see Figure 2.4)

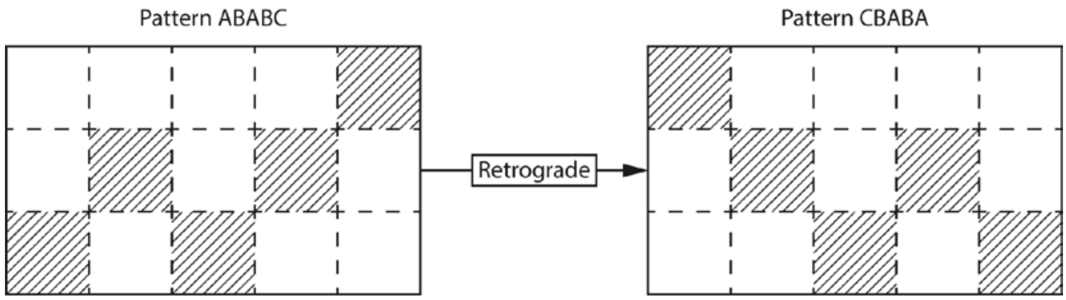


Figure 2.4. Pattern ABABC and its retrograde CBABA.

Third, all patterns were required to produce three unique transformations (inverse, retrograde and retrograde inverse). Any inherent symmetry in a pattern meant that it could only provide a single unique version under transformation. For example, the pattern ABCBA produces the pattern CBABC under inverse transformation. However, its retrogression is structurally identical to the original pattern, and its retrograde inversion is structurally identical to its inversion (see Figure 2.5), making them redundant for the purposes of the transformation task.

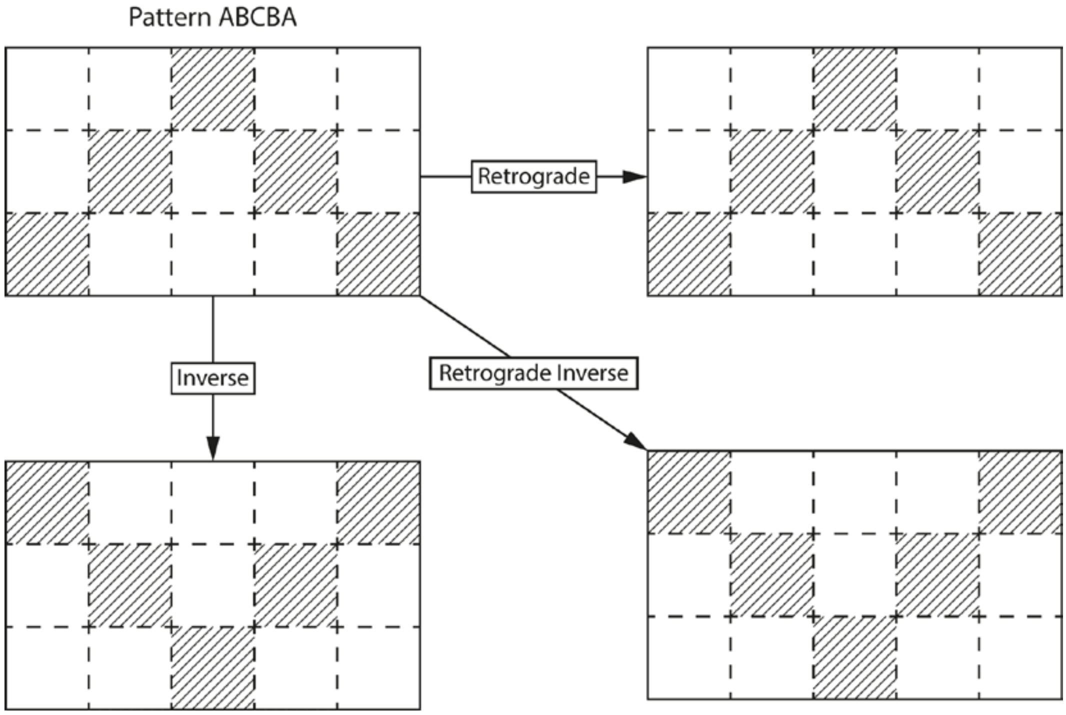


Figure 2.5. Pattern ABCBA and its transformations.

Fourth, each pattern was required to end on a different value to the one it started on. This criterion was set in order to ensure that responses were not based on strategies that did not require the mental transformation of the entire pattern. It was anticipated that when presenting two unrelated patterns where one begins and ends on the same value and the other begins and ends on different values, a decision about their relatedness could be based on simply recognising this fact.

After applying these criteria to the complete pattern set, a pool of 27 unique, transformationally distinct patterns remained. There were four *transformational variants* of each pattern (the original pattern plus its three transformations: inverse, retrograde and retrograde inverse) making a total of 108 patterns. Fifteen transformationally distinct patterns were randomly selected for use in the experiments as standard and related target stimuli (see Table 2.4). For each of the 15 selected patterns, one of its transformational variants was chosen as the standard with the criterion that an equal proportion of standard patterns started on each of the three values A, B and C (so that 5 started on A, 5 started on B and 5 on C). Next, another transformationally distinct pattern was randomly selected from those remaining in the pool for use as an example in the experimental training. It should be noted that at this point all of the retrograde inversion variants of the 15 standards and example pattern could be discarded, as this transformation condition was not included in the experimental design.

Table 2.4

Standard and related target patterns used in Experiments 2, 3, 4 and 5

Standard	Related Target	
	Inverse	Retrograde
AABBC	CCBBA	CBBAA
AACAB	CCACB	BACAA
ABAAC	CBCCA	CAABA
ABCAB	CBACB	BACBA
ABCCC	CBAAA	CCCBA
BACBC	BCABA	CBCAB
BBAAC	BBCCA	CAABB
BBACA	BBCAC	ACABB
BBBAC	BBBCA	CABBB
BCBAA	BABCC	AABCB
CABAA	ACBCC	AABAC
CABAB	ACBCB	BABAC
CABBA	ACBBC	ABBAC
CACAB	ACACB	BACAC
CACBA	ACABC	ABCAC

The remaining 11 unused transformationally distinct patterns and their variants (a total of 44 patterns) were pooled for use as unrelated target stimuli. Selection of unrelated target patterns in the experiment could not be completely random - it was anticipated that unrelated target patterns may be easily detected if they began on a different value to the related target, as this would allow participants to base their judgments on the first event of the pattern alone, rather than on the pattern as a whole. Therefore, unrelated target patterns were further pooled according to the first event value, and the following rule was applied to their selection: in any given trial the unrelated target pattern was required to begin on the same value as the corresponding related target pattern. To illustrate, in the retrograde condition a related target pattern always began with the same value as the one that the standard ended with. For example, the standard pattern ABABC ended on value C, thus its retrograde CBABA also began on value C. In this case, an unrelated target was randomly selected from the pool of patterns that began on value C. Correspondingly, in the inverse condition a related target pattern always began with the inverse value of the one that the standard began with. For example, the standard pattern ABABC began on value A, thus its inverse CBCBA began on value C. In this case, an unrelated target was randomly selected from the pool of patterns that began on value C. See Table 2.5 for all unrelated target patterns.

Table 2.5

Unrelated target patterns used In Experiments 2, 3, 4 and 5 (grouped by starting value)

Group A	Group B	Group C
AAACB	BAAAC	CAAAB
AACBB	BAABC	CAACB
AACBC	BAACA	CAACB
AACCB	BAACC	CACCB
ABACC	BABBC	CBAAB
ABBCB	BABCA	CBABB
ABCBB	BACCA	CBBAB
ABCCB	BACCC	CBCAA
ACAAB	BBABC	CCAAB
ACBAB	BBACC	CCABA
ACCAB	BBCAA	CCABB
ACCCB	BBCBA	CCCAB
	BCAAA	
	BCAAC	
	BCBAC	
	BCBBA	
	BCCAA	
	BCCAC	
	BCCBA	
	BCCCA	

2.4.2 Auditory stimuli

The auditory stimuli used in all experiments were pitch patterns that were monophonic (only one pitch was heard at any given time) and isochronous. The majority of previous research on the recognition of melodic pitch patterns has used stimuli composed from the diatonic scale (e.g. Reiner, 2011; Trainor, Desjardins, & Rockel, 1999; Trehub, Bull, & Thorpe, 1984), which is the scale most commonly used in Western music and involves whole-tones and semitones of the 12-note equal temperament scale. However, tonality has been shown to influence listeners' judgments of pitch interval distance (see Shepard, 1982, p.315). Though steps of the diatonic scale may be different in interval size, they are equal in terms of tonal distance. As such, (depending on the tonal context) a semitone interval may be perceived as being equivalent in size to a whole-tone interval (Shepard, 1982). This phenomenon would be problematic for the purposes of the present research, as perceptions of pitch interval must be compared directly with perceptions of visual distance. Therefore, an alternative tuning system to the 12-note chromatic scale (on which the diatonic scale is based) was adopted for the present stimuli. By dividing the octave into 5 equal intervals, the influence of tonality is avoided to some extent. This tuning has been used in some recent studies (Overath et al., 2007; Stewart, Overath, Warren, Foxton, & Griffiths, 2008).

The fundamental frequencies of tones used in the experiments were determined using the following formula for equally-tempered scale systems (White, 2005):

$$F_n = F_a\left(\sqrt[N]{I}\right)^{(n-a)}$$

where F_n is the frequency to be calculated, F_a is the frequency of a reference pitch, I is the bounding interval, N is the number of notes within the bounding interval, and n and a refer to numbers assigned to the to-be-calculated frequency and the reference frequency respectively. In order to calculate F_n the following constants were set. In the case of a 5-note chromatic scale the bounding interval is an octave (a doubling of frequency) hence I is 2, and N is 5. The reference pitch frequency (F_a) was set at 260.00 Hz. This is in keeping with previous research (e.g. Dowling, 1972; Edworthy, 1985; Eerola, Jarvinen, Louhivuori, & Toiviainen, 2001; Freedman, 1999; Thorpe, Ockelford, & Aksentijevic, 2012) and meant that pitch patterns were located within a pitch range employed in everyday musical settings (concert pitch middle C, or C_4, is designated the frequency 261.63 Hz). The reference frequency was the first note in the scale, thus a is 1. Finally, the remaining variable, n, depended on which step of the scale was being calculated. As the range of values in patterns for the present experiments was set to three, n is simply 1, 2 or 3. With all of these values set, the calculations produced the frequencies displayed in Table 2.6. The intervals between tones were equivalent to approximately 240 cents.

Table 2.6

Calculation of frequencies in Hz (F_n) for tones used in the research

Value	F_a	I	N	n	a	Equation	F_n
A	260	2	5	1	1	$260(\sqrt[5]{2})^{(1-1)}$	260.00
B	260	2	5	2	1	$260(\sqrt[5]{2})^{(2-1)}$	298.66
C	260	2	5	3	1	$260(\sqrt[5]{2})^{(3-1)}$	343.07

Auditory stimuli consisted of triangle waveforms that included the fundamental frequency and a profile of corresponding odd harmonics of the series (as employed by Widmann, Kujala, Tervaniemi, Kujala, & Schröger, 2004) (see Figure 2.6). Although pure tones have been used in some of the previous literature, there were two reasons why they were considered unsuitable for the present research. Firstly, they sound artificial and are rarely heard in the natural environment, which means they have less ecological validity when compared with complex tones. Secondly, pure tones are informationally poor - it has been shown that the pitch of complex tones is discriminated more accurately than the pitch of pure tones (see Sidtis, 1980) - and it follows that pitch patterns produced with complex tones are encoded more efficiently. The triangle wave was selected over other waveforms (e.g. square, saw-tooth) as it was relatively pleasant to the ear but also had a neutral timbre that was not relatable to a familiar instrumental sound.

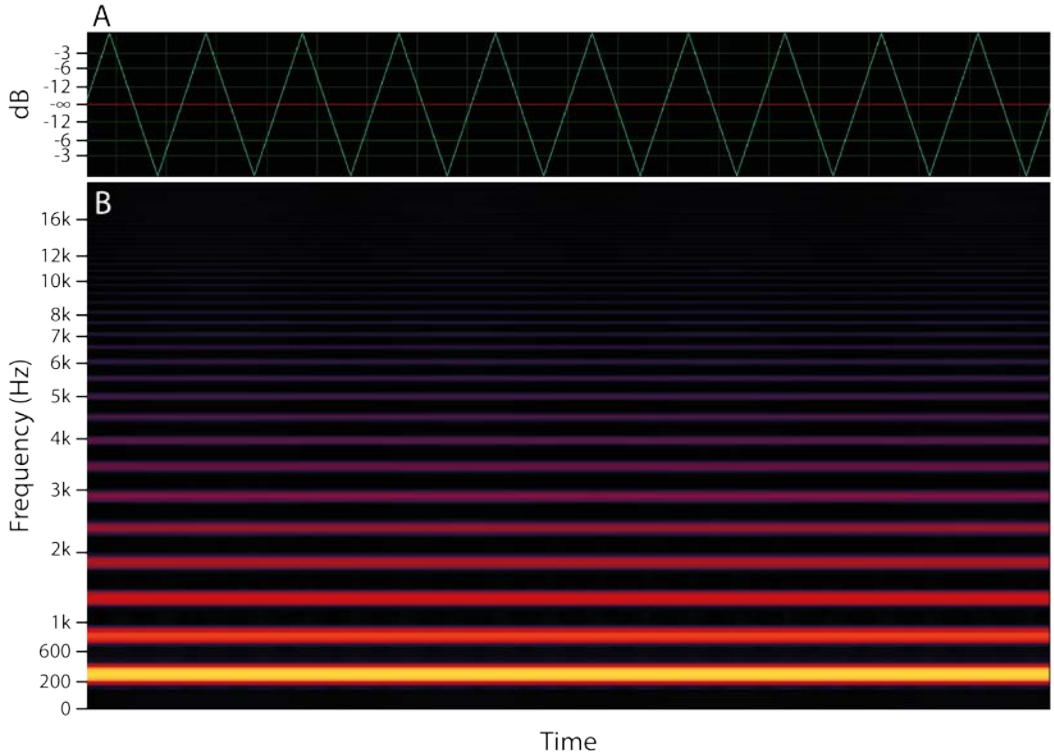


Figure 2.6. Sample of a stimulus tone with f_0 260 Hz (duration: 10 cycles of the sound wave [approximately 38.46ms]). (A) the waveform; (B) the time-frequency plot – visible are the f_0 at 260 Hz and corresponding odd harmonics (third harmonic 780 Hz, fifth harmonic 1,300 Hz, seventh harmonic 1,820 Hz, ninth harmonic 2,340 Hz, etc.).

In Experiments 2 and 3 tone durations were 500ms and there were no inter-stimulus intervals (ISI) between tones in a sequence (see Figure 2.7A). This is in keeping with Dowling (1972) who showed that participants could successfully identify melodic transformations at this presentation rate. Also, Balch and Muscatelli (1986) showed that in unimodal and cross-modal recognition experiments of pitch and visual contour, there was a noticeable drop-off in performance accuracy with faster presentation rates. In Experiments 4 and 5 it was necessary to shorten tone durations to 350ms whilst maintaining the same inter-onset interval (IOI) of 500ms between tones (a detailed explanation of why this step was taken can be found in the next section on visual stimuli). This created a 150ms ISI between the offset and the onset of tones in a sequence (see Figure 2.7B). A 10ms linear amplitude ramp was applied to the onset and offset of all tones.

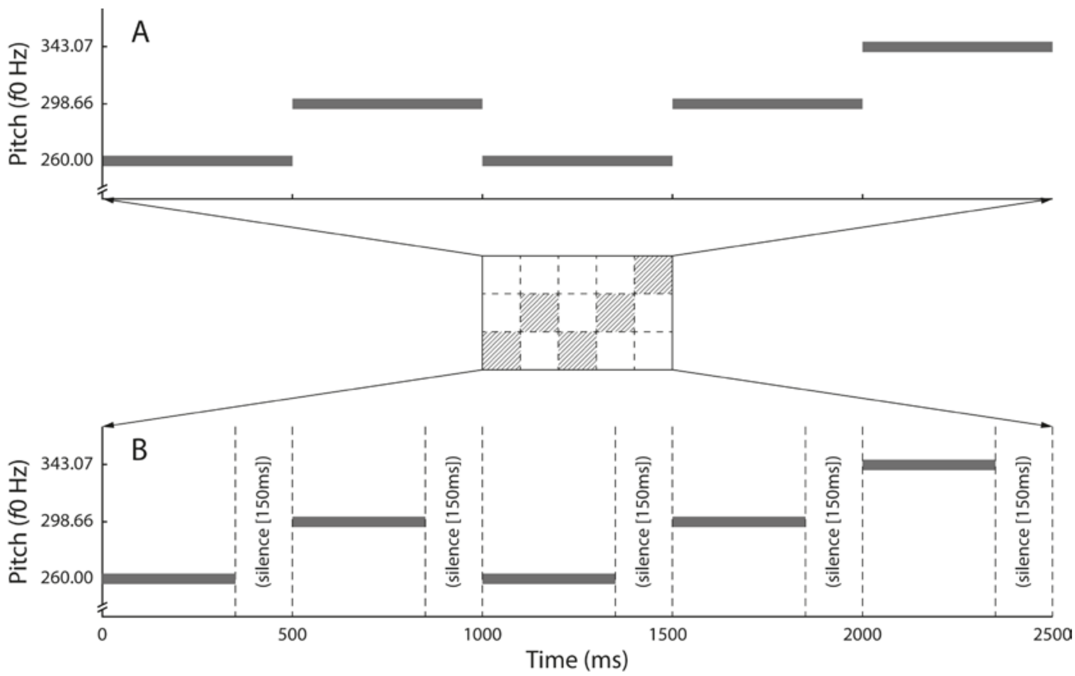


Figure 2.7. Auditory stimuli generated from the pattern ABABC, as presented in Experiments 2 and 3 (A), and Experiments 4 and 5 (B).

Tones were generated using NCH Tone Generator version 3.02 (NCH Software), and then edited using WavePad Sound Editor Masters Edition version 5.02 (NCH Software). WavePad was also used to arrange tones into sequences which were digitally recorded as .wav file type, sample size 16 bit, sample rate 44 kHz, format PCM uncompressed, mono.

2.4.3 Visual stimuli

Key to the premise of the research is the analogy between auditory pitch patterns and visuo-spatial patterns, and therefore what form a visual stimulus should take in order to be perceptually equivalent to a pitch pattern. As discussed above, a variety of approaches to presenting analogous visual stimuli have been taken in previous research. Inconsistency in the way in which analogous visual patterns have been presented in the existing literature meant that a number of issues were taken into consideration when designing the visual stimuli used in the present research.

As the present research sought to treat auditory and visual patterns as equitably as possible, all visual stimuli were presented sequentially. Pitch patterns are temporal, and evolve over time. In order to perceive a temporal pattern as an integrated whole, all of the elements of the sequence must be held in memory. The perception of a visual pattern that is presented all at once would not place the same demand on cognitive resources.

The visual stimulus was a sequence of discrete objects presented at different spatial positions on a computer screen. Objects at different vertical heights represented different auditory pitches, with higher pitches corresponding to higher spatial positions. The form of individual objects in analogous visual stimuli does not appear to have been taken into serious consideration in previous research, despite evidence to suggest that different sounds are associated with visual objects of different size and shape (see Spence, 2011, for a review). It was reasoned that an object that clearly marks its spatial position on the vertical axis would be suitable for the present experiments. With this in mind, a black bar (or linear segment) with rounded ends was produced (see Figure 2.8) that measured

2.72 by 0.55 cm, so they subtended a visual angle of 2.23 by 0.45˚ at a viewing distance of 70 cm.[14]

Experiments 2 and 3 used visual stimuli that were presented horizontally from left to right. Sequences were presented on the screen in such a way that there was no gap separating each successive linear segment on the horizontal axis. The width of a complete sequence measured 13.65 cm, which subtended a visual angle of 11.15° at a viewing distance of 70 cm. Experiments 4 and 5 used visual stimuli where all segments appeared at different locations on the central vertical axis only.

The interval distance (vertical height) between segments also needed to be considered. For pitch patterns, the pitch distance between tones was dictated by the 5-note equal temperament tuning. What the equivalent would be in terms of visual height has not been openly considered in previous studies, and yet is fundamentally important to the development of analogous auditory and visual stimuli. Experiment 1, reported in Chapter 3, was therefore carried out in order to address this issue. To summarise briefly, participants were asked to map different pitch intervals from the 5-note equal temperament scale employed in the present research onto a computer screen. Based on the findings of this experiment, the equivalent of one pitch interval was set to a visual angle of 1.35˚, which, at a viewing distance of 70 cm, translated into 1.65 cm. This in turn meant that the height of all complete visual sequences from the lowest event to the highest was 3.85 cm, which subtended a visual angle of 3.15˚ at a viewing distance of 70 cm.

[14] Visual angles are reported in degrees and decimals rather than in degrees, minutes and seconds. This is in keeping with previous research (e.g. Pomerantz & Portillo, 2011).

Finally, the display durations and IOIs for each segment of the sequence were set to match those of the auditory stimuli. For Experiments 2 and 3 the duration of presentation of each segment was 500ms with no ISI between events (see Figure 2.8A). However, when visual stimuli were presented centrally only, successive segments presented at the same vertical position were perceptually impossible to segregate. Therefore, in Experiments 4 and 5 display durations for each segment were reduced to 350ms, thereby creating a 150ms ISI after each event (see Figure 2.8B). This duration was selected as it produced an easily perceivable segregation between two successive segments presented at the same height. In all experiments the IOI between successive segments was 500ms.

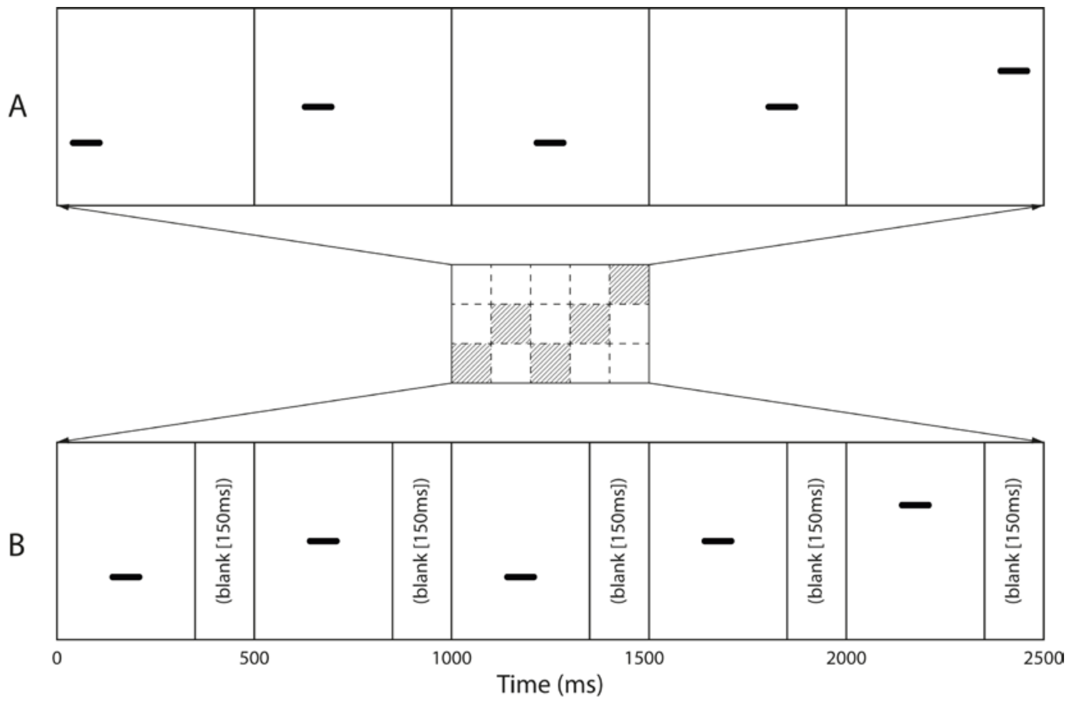


Figure 2.8. Visual stimuli generated from the pattern ABABC, as presented in Experiments 2 and 3 (A), and Experiments 4 and 5 (B).

Image stills were produced with Adobe Illustrator CS5 version 15.1.0 (Adobe Systems Incorporated) and exported in .jpeg file format. The image stills were animated using Final Cut Pro and exported as QuickTime files before being converted to WMV format using MPEG Streamclip version 1.9.2. In all experiments the frame of the video file was aligned to the centre of the display screen. This meant that the middle value (the vertical height equivalent to 298.66 Hz) was aligned to the centre of the screen on the vertical axis. In Experiments 2 and 3 the third event of the sequence (half way between the beginning and the end) was aligned to the centre of the screen on the horizontal axis.

2.5 Design

In all recognition experiments there were three independent variables, each with two levels: modality (auditory [A], visual [V] in unimodal trials [Experiments 2 and 4]/ auditory-visual [AV], visual-auditory [VA] in cross-modal trials [Experiments 3 and 5]/ auditory standard [AS], visual standard [VS] in hybrid trials [Experiment 6]), relatedness (related target, unrelated target), and transformation (inverse, retrograde). The transformation factor was embedded in the related level of the relatedness factor. As a result, there were six experimental conditions. Each standard pattern was presented once in related conditions and twice in unrelated conditions, hence the proportion of trials per condition was as follows: 1) A/AV/AS, related, retrograde = 12.5%; 2) A/AV/AS, related, inverse = 12.5%; 3) A/AV/AS, unrelated = 25%; 4) V/VA/VS, related, retrograde = 12.5%; 5) V/VA/VS, related, inverse = 12.5%; 6) V/VA/VS, unrelated = 25%. Fifteen unique standard patterns were used in Experiments 2, 3, 4 and 5, making a total of 120 trials per experimental session. Ten unique standard patterns were used in Experiment 6, making a total of 80 trials.

Experiments 2, 3, 5 and 6 employed a within-subjects factorial design with each participant being tested in all experimental conditions. The block structure for the experimental session is displayed in Figure 2.9. The session was divided into two blocks containing equal proportions of trials from each modality condition and equal proportions of related and unrelated trials. Participants only had to recognise one type of transformation per block, therefore one block contained all the related inverse trials, and the other block contained all related retrograde trials. Each block was further sub-divided by modality into sub-blocks containing equal proportions of related and unrelated trials. The order of blocks was counterbalanced between participants. The order of sub-blocks within each block was randomised. The presentation order of trials within each sub-block was randomised.

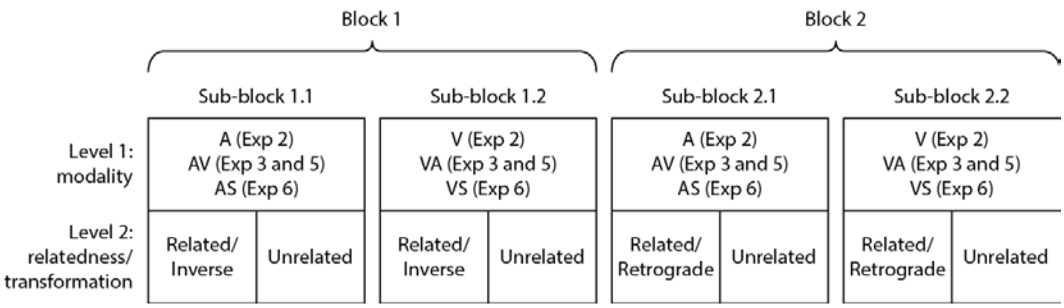


Figure 2.9. Block structure for Experiments 2, 3 5 and 6.

For Experiment 4 a mixed factorial design was employed in which each participant was tested in all experimental conditions except for modality (auditory, visual), which was made a between-subjects factor. The block structure for the experimental session is displayed in Figure 2.10. The assignment of participants to experimental sessions of different modality conditions was counterbalanced. Each experimental session consisted of two blocks with equal proportion related and unrelated trials. Participants only had to recognise one type of transformation per block, therefore one block contained all the related inverse trials, and the other block contained all related retrograde trials. The order of blocks was counterbalanced between participants. The presentation order of trials within each block was randomised.

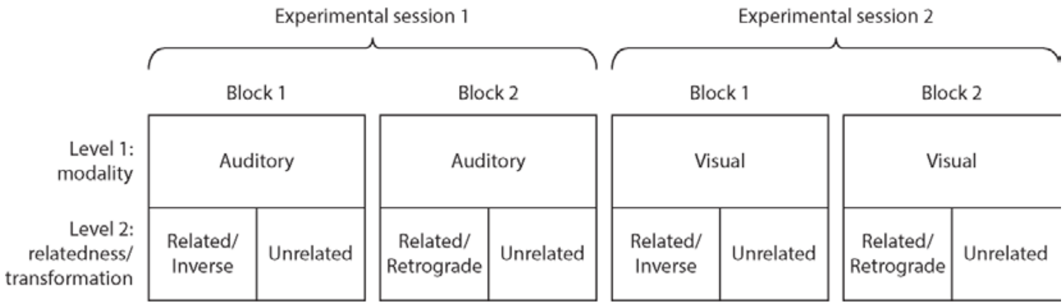


Figure 2.10. Block structure for Experiment 4.

2.6 Apparatus

All the listed experiments were carried out in a sound-attenuated room. Background noise level was minimal, not exceeding 30 dB SPL, and ambient luminance was kept at a constant low level in order to provide a non-distracting experimental environment.

Delivery of experimental trials, stimulus generation and response recording were controlled by a Dell OPTIPLEX 760 PC. Sound stimuli were generated by a SoundMAX HD Audio soundcard and delivered binaurally using Shure SRH440 headphones. Intensity of the sound stimuli was approximately 60 dB SPL, examined using an artificial ear and an Adastra analog sound level meter (Model 952.422, slow response). Visual stimuli were generated by a NVIDIA Quadro FX 580 graphics card and presented on a Dell Trinitron P1130 CRT monitor with screen dimensions 40.50cm by 30.00cm. Contrast and brightness were adjusted to a comfortable level.

Experimental procedure, stimuli sequencing and response recording was executed using E-Prime version 2.0 (Psychology Software Tools, Pittsburgh, PA). All responses were made on a PST Serial Response Box Model 200A.

2.7 Procedure[15]

Participants were tested individually, and were not permitted to take part in more than one of the experiments reported in this thesis. On arrival, they were supplied with a briefing sheet and consent form, and asked to complete a short 'Demographic and Music Background Questionnaire'. This collected demographic data such as age, gender and handedness, and also asked the participant to self-report any hearing problems that might interfere with their ability to perform the task. It also collected musical experience data such as months/years of training, ability to read music notation, instruments played and time since last performance. Once this was completed participants were brought to the experimental room. Care was taken to ensure that the participant was comfortably seated with their eye level approximately 70cm from the centre of the computer screen (see Figure 2.11).

[15] Examples of supporting documents (i.e. briefing, consent forms, debriefing, questionnaires) and instructions given to participants in all experiments can be seen in Appendix I.

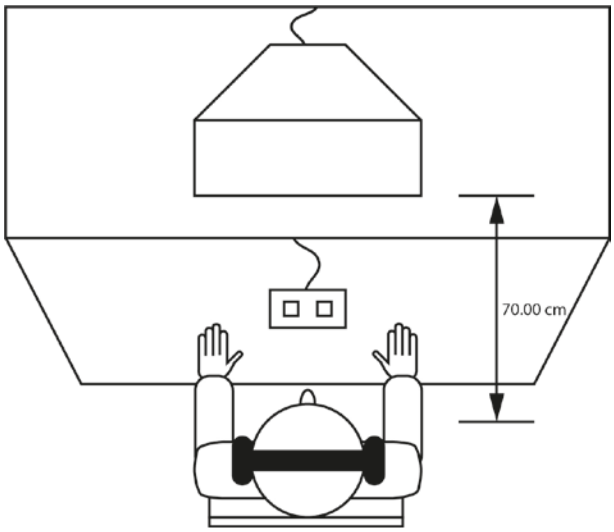


Figure 2.11. The experimental apparatus. Participants were seated in front of the computer screen and wore a pair of headphones. Responses were made on the response box placed in front of them using the index and middle fingers of their dominant hand.

Training for the present experiments was of great importance due to the anticipated high task difficulty (see Chapter 1, Section 1.5.3.3). Though participants with some musical experience might have been familiar with the concept of comparing melodic permutations and relating them to visual representations, for most participants this would have been a completely novel task. Extensive training was therefore delivered by the experimenter on the computer at the beginning of the experiment, before each main block and before each sub-block, in order to ensure participants were familiar with the concepts involved in performing the task. The following paragraphs will describe the training procedure in full. Note that in Experiment 4 participants only took part in one level of modality, therefore the instructions were adapted accordingly.

At the beginning of the experiment, participants were told that they would be presented with short "melody-like" patterns on the headphones and/or with short sequential patterns of objects on the computer screen. Participants were then presented with auditory and visual example patterns. In Experiment 4 participants were only presented with an auditory or a visual pattern, depending on the modality condition to which they had been assigned. Participants were then informed that they would need to compare different patterns with each other to determine whether they are related or unrelated under different types of transformation.

It was unnecessary to inform participants about both types of transformation before the first main block, so they were only informed about the relevant transformation for the block in which they were about to take part. Inverse transformations were described both as ‘upside down’ versions of a standard pattern, and as inversions of a pattern where an ‘up’ becomes a ‘down’ and a ‘down’ becomes an ‘up’. Retrograde transformations were described both as ‘backwards’ versions of a standard pattern, and as patterns presented in reverse order. The description was illustrated by presenting again the previously shown example pattern, this time followed by its relevant transformation.

Finally, before each sub-block, participants were informed of the modality condition that patterns were to be presented in (except in Experiment 4 where this was not necessary) and took part in six un-timed example trials. At this point they were advised how to make their responses in order for them to get some practice while the experimenter was still present to supervise. Participants were instructed to indicate whether they thought the target pattern was related to the standard or was unrelated by pressing one of two buttons on the response box provided. They were told to use their index and middle fingers of their dominant hand to respond. In 50% of experimental sessions ‘related’ responses (indicating a related target) were allocated to the left button and ‘unrelated’ responses (indicating an unrelated target) were allocated to the right button of the response box. In the other 50% 'related' responses were allocated to the right button and 'unrelated' responses were allocated to the left button of the response box. The example stimuli were presented in the modality that corresponded to the conditions of the sub-block. After each response feedback was provided. Participants were encouraged to ask questions throughout, and any issues were discussed.

Once the training was finished participants were informed explicitly of what would be happening on each trial of the experiment (see Figure 2.12). First, they were told that at the beginning of each trial they would see a fixation cross and hear a warning beep. The fixation cross measured 0.45 x 0.45° and was presented at the centre of the screen. The warning beep was a pure tone, frequency 298.66 Hz, with a duration of 200ms and linear amplitude ramps of 3ms onset/6ms offset. This frequency was chosen because it was the same as the f_0 of the middle pitch of the auditory stimuli. The warning beep thereby acted to draw participants' attention to the centre of the pitch range of the stimuli, much as the fixation cross acted to draw attention to the centre of the orientation of the visual stimuli. Second, they were told that in each trial they would be presented with the standard pattern followed (after a short pause) by the target pattern. The standard stimulus was presented 1000ms after the onset of the fixation-cross and warning-beep. On the offset of the standard there was a 2000ms ISI in which the fixation point reappeared before the target stimulus was presented. Third, they were told that once the target had finished, there was a time limit for responses before the computer moved on automatically to the next trial. From the target stimulus offset the screen remained blank and participants were given up to 5000ms to respond –

if a response was made before the time limit was reached the experiment immediately moved onto the next trial.

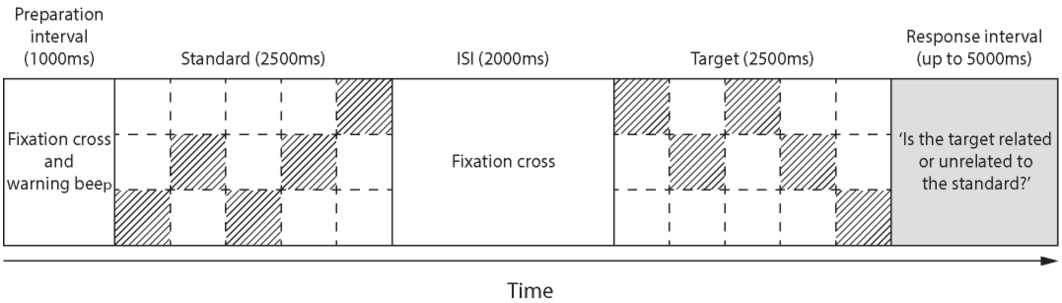


Figure 2.12. Time course of an experimental trial in Experiments 2, 3, 4, 5 and 6. The target is an inverse transformation of the standard. Targets could also be retrograde transformations or unrelated to the standard.

In addition to the above description, some guidelines for making responses were offered. The intention was not to bias the participants' responses, but rather to encourage them to focus on the contour of the pattern when performing the task, and also to ensure some level of agreement between participants in terms of how they approached the task. Participants were advised that decisions should be based on the whole pattern, not just features of the patterns such as the beginning or the end. It was made clear that responses must be made after the target had finished and that responses made before this would not be registered. This requirement ensured that the target as a whole would be taken into account when giving responses. RT was therefore measured from the offset of the target stimulus. Participants were told that though RT was being measured, it was more important to be as accurate as possible. They were told this in an attempt to ensure that the error rate was manageable. Finally, they were informed that the target was equally likely to be related or unrelated to the standard. When the participant was ready to start, the experimenter left the room and did not return until the end of the sub-block. Before starting the experimental trials, participants took part in 6 practice trials. Feedback was provided for responses to practice trials, but not for responses to the experimental trials.

Once the experiment had finished comments were recorded from the participants relating to their experience. The experimenter asked them how they approached performing the task, whether they were aware of employing any strategies, and whether they felt they performed differently across the different conditions. Finally, a debriefing was administered in which participants were informed of the purposes of the experiment and supplied with contact information for the research team should they need to be contacted. The entire experimental session took on average between 50 and 55 minutes to complete (between 25 and 30 minutes in Experiment 4).

2.8 Data analysis

Individual data sets were automatically compiled by E-Prime in .edat format at the end of each experimental session. Once all the data for a particular experiment were collected, individual data sets were merged and then transferred to SPSS (versions 19 and 20) on which all of the analysis was performed.

For the preliminary analysis, data were converted into aggregated form and mean PE and RT scores were calculated across conditions. An arcsine transformation was performed on PE scores (Howell, 2002, p.347) and a natural-logarithmic transformation was performed on RT to account for a positively skewed distribution of the variable (Howell, 2002, p.344). The log transformation was applied to the raw RT data, before aggregation. All inferential statistical analyses were performed on these arcsine and log-transformed data, though all descriptive statistics were reported in the units of the untransformed variables. Similarly, all graphs displayed the untransformed data and error bars were corrected for within-subject variability (Loftus & Masson, 1994) in order to be more informative of the effects.

Analyses of variance (ANOVAs) were run to examine the data. Within-subjects ANOVAs were run on the data collected from Experiments 2, 3, 5 and 6, and mixed-design ANOVAs were run on the data collected from Experiment 4. In each case, an ANOVA was carried out firstly to examine the effects of relatedness. Further ANOVAs were carried out which focussed on examining the effects of modality and transformation on related trials only. Where the assumption of sphericity was violated, Greenhouse-Geisser correction for degrees of freedom was applied (Howell, 2002, p.486). Corrected degrees of freedom were reported except in cases in which uncorrected and corrected F values were identical. Planned and post-hoc pairwise comparisons were carried out on the estimated marginal means to test hypotheses and explore any unpredicted significant findings. It was unnecessary to apply a Bonferroni correction to the mean differences between conditions for post-hoc comparisons as there were never more than two levels of a variable (Howell, 2002, p.384).

A secondary analysis utilising signal detection theory (SDT; Wickens, 2001) was run in order to examine more closely the structure of error data and factors that influence error. SDT can be applied whenever two possible stimulus types must be discriminated and was historically applied in studies of perception, where participants discriminated between *signals* (stimuli) and *noise* (no stimuli). It has since been applied to many other areas including recognition memory, in which the signal corresponds to old items (related targets) and the noise corresponds to new items (unrelated targets). The underlying premise is that when making a yes/no decision in a task in which the signal is not easily separated from the noise, there will be some level of uncertainty involved. In this case, statistics based on SDT measures give arguably a better representation of a participant's sensitivity to the signal than PE alone (Wickens, 2001). The experimental task in the present research was very difficult for some participants. Thus, it is possible that the observed patterns of PE were influenced not only by the sensory process of interest but also by a decision process. SDT provides a solution to this by offering a method of decomposing mental processes into sensory and decision subprocesses.

In SDT, a participants' sensitivity to the signal can be operationalised in terms of hits (the proportion of responses that are correct identifications of the signals) and false alarms (the proportion of responses that incorrectly indicated a presence of the signal). A number of measures can be quantified from these proportions – d-prime (*d'*) was deemed most appropriate for use in the present research because of its application in many contemporary studies investigating the recognition of pitch patterns (e.g. Williamson, Baddeley, & Hitch, 2010; Wong et al., 2012). When using *d'* as a measure, a value of 0 indicates an inability to distinguish signals from noise, whilst more positive values indicate a correspondingly greater ability to distinguish signals from noise.

In addition to quantifying a measure of sensitivity to the signal, a measure describing response bias was also employed. Response bias refers to the participants' general tendency to respond *yes* (i.e. related) or *no* (i.e. unrelated), and is determined by the location of the decision criterion (Wickens, 2001). The decision criterion represents the location of a participant's decision profile with respect to theoretical signal and noise distributions. For the purposes of the current research, response bias was measured with *c*, which is defined as the distance between the criterion and the neutral point, where neither response is favoured. Negative values of *c* signify a bias toward the *yes* response – a 'liberal' response strategy. Positive values of *c* signify a bias toward the *no* response – a 'conservative' response strategy. Bias is independent of sensitivity (Wickens, 2001) and can thus reveal effects of the experimental conditions that may not be revealed by a measure of sensitivity alone.

In order to carry out the SDT analyses, error data were first converted to hits and false-alarms. d' and *c* measures were then calculated using a log-linear approach in order to account for instances in which hit and false-alarm rates were equal to 0 or 1 (Stanislaw & Todorov, 1999, p.144). The mathematical formulae for d' and *c* measures were applied in SPSS using syntax published in Stanislaw and Todorov (1999). Within-subjects and mixed-design ANOVAs were then run on these measures in order to examine the effects of modality and transformation.

Finally, based on self-reported data collected from the questionnaires, participants were placed into one of two groups: some training, or no training. Participants were allocated to the some training group if they had received at least 6 months training on any instrument or voice. The ANOVAs examining the effects of modality and transformation were then repeated with music training included as a between-subjects factor.

2.9 Participant selection and treatment

Participants who took part in this research were recruited from two main sources. The main source of recruitment was via the Department of Psychology's on-line participant recruitment system. This system facilitates the participation of undergraduate psychology students who in exchange receive course credit. Students recruited this way needed to collect 12 credits as a compulsory requirement for completion of the Experiencing Psychological Research module. They received one credit for every hour's worth of participation. The secondary source of recruitment was via poster, email and word of mouth. This method included the recruitment of postgraduate students, and members of the public. In these instances participants received a £10 gift voucher as reimbursement for their involvement. A majority of participants were female (74%) and age ranged from 18 to 73 years (M = 24.01, SD = 9.63).

On arrival participants were greeted by the experimenter and given a written briefing. Once they had read this and the experimenter had confirmed that they understood the task, informed consent was collected. Before each session, participants were also given a questionnaire to collect demographic information including gender, age, handedness and first spoken language. They were also asked whether they suffered from any hearing problems. Details about their musical experience were also collected, including the number of years and nature of any formal or informal education, instruments played and most recent performance.

The experimenter was in the testing room with the participant in order to deliver the training, but only the participant was present throughout the practice and experimental trials. At the end of each block participants were asked how they found the trials, and at the end of the experiment they were asked to comment on their impressions and experiences. Finally, participants were handed a written debrief and informed of the aims of the research.

2.10 Ethics considerations

Before any of the reported experiments were carried out a detailed account of the methods to be used was submitted for approval to the Department of Psychology, University of Roehampton, Ethics Committee. At the University of Roehampton, responsibility for consideration of ethical issues is devolved to the Department, which must consider the application against the University's Ethics and the guidelines set out by the BPS. Once approved, the Head of Department must confirm the assessment of the application before sending to the Ethics Administrator. Establishment and monitoring of Departmental procedures is undertaken by the University's Ethics Committee, which meets three times a year. All aspects of the project were approved and communicated by the Ethics Administrator on 7th September, 2011.

All data were stored on a password protected computer in a secure office on the University grounds. Access to this computer was exclusive to the investigator. Hard copy questionnaires and behavioural data were stored securely in a separate locked filing cabinet. In compliance with participant confidentiality and anonymity, data was linked with participant demographics by participant ID number only. Consent forms were not linked to ID number and were kept in a locked cabinet, separately from questionnaire and behavioural data. On project completion, all data will be stored in a secure location for a period of at least 10 years from the date of any publication which is based upon it.

2.11 Summary of experiments

Table 2.7

Summary of experiments

Experiment	Paradigm	Trials	Modality condition	Modality of the stimulus		Transposed target	Visual stimuli
				Standard/Prime	Target		
1	Cross-modal psychophysical scaling	Cross-modal	n/a	n/a	n/a	n/a	n/a
2	Short-term recognition	Unimodal	Within-subjects	Auditory or Visual	Auditory or Visual	No	2½-D
3	Short-term recognition	Cross-modal	Within-subjects	Auditory or Visual	Auditory or Visual	n/a	2½-D
4	Short-term recognition	Unimodal	Between-subjects	Auditory or Visual	Auditory or Visual	No	1½-D
5	Short-term recognition	Cross-modal	Within-subjects	Auditory or Visual	Auditory or Visual	n/a	1½-D
6	Short-term recognition	Hybrid	Within-subjects	Auditory or Visual	Auditory and Visual	Yes	1½-D
7	Structural priming	Unimodal	n/a	Auditory	Auditory	Yes	n/a
8	Structural priming	Cross-modal	n/a	Auditory	Visual	n/a	1½-D

Chapter 3: Finding a common structural metric between auditory pitch and visual space

3.1 Experiment 1: Introduction

The present chapter reports on an experiment that was carried out to address the question of whether there is a common structural metric between auditory pitch and visual space. Ultimately, the results of this experiment determined the design of the visuo-spatial stimuli used in all subsequent experiments.

3.2 Background and rationale

Although there is evidence for both the explicit and the implicit mapping of pitch height onto visual height (e.g. Evans & Treisman, 2010; Lidji et al., 2007; Rusconi et al., 2006), the potential for a common structural metric across both modalities remains (to the author's knowledge) unexamined. The present research set out to treat sound and vision as equitably as possible, and, should a common structural metric exist between modalities, it follows that this should be taken into account when designing analogous auditory and visual stimuli. This issue has not been considered in the previous literature addressing the cross-modal cognition of auditory pitch and visual contour.

In order to examine and measure participants' potential common structural metric for pitch height and visual height, a cross-modal scaling task adopted by Mudd (1963) was adapted for replication in Experiment 1. Mudd's original experiment was carried out to explore the potential spatial stereotypes relating to the frequency of pure tones. Participants took part in a number of trials in which they heard a reference tone of a particular frequency that was followed by a comparison tone of a different frequency. The reference tone was represented by a peg positioned at the centre of a pegboard, and in each trial their task was to reposition the peg on the board in order to represent the comparison tone. Minimal instructions were given to participants with regard to how they should complete the task and, according to Mudd, decisions informing the repositioning of the peg were entirely arbitrary. It was found that the orientation of responses was remarkably consistent across participants (see Figure 3.1). On the vertical axis, higher frequency comparison tones were represented by pegs repositioned above the reference peg and lower frequency comparison tones were represented by pegs repositioned below the reference peg. In addition to this, on the horizontal axis higher frequency tones were represented by pegs repositioned to the right of the reference peg and lower frequency tones were represented by pegs repositioned to the left of the reference peg.

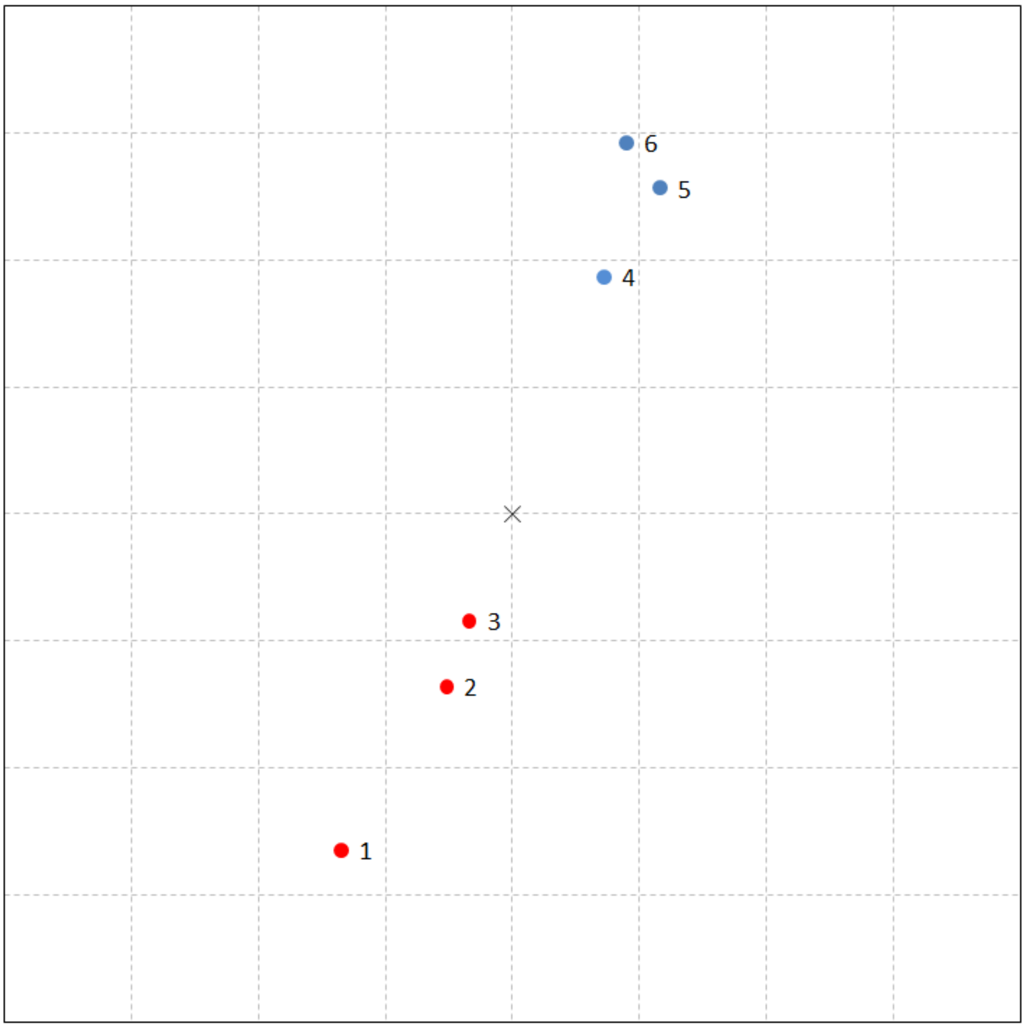


Figure 3.1. Illustration of the data reported by Mudd (1963). Plots represent the positioning of a reference peg made in response to different comparison tones (1 = 222 Hz, 2 = 402 Hz, 3 = 748 Hz, 4 = 2574 Hz, 5 = 4787 Hz, 6 = 8861 Hz). The reference peg was initially placed at the centre of the pegboard (40 cm x 40 cm) and represented a reference tone (1391 Hz) that preceded each comparison tone.

As sounds were presented sequentially, it could be argued the horizontal deflection reflects the temporal dimension of time (see Casasanto, 2010). However, this explanation does not account for why higher tones were represented by pegs repositioned to the right while lower tones were represented by pegs repositioned to the left, despite them both being presented after the reference tone. Rather, it appears that the spatial mapping of pitch requires at least two dimensions. Recent experiments have indeed confirmed that pitch may be mapped onto a horizontal dimension of space as well as the vertical dimension (Evans & Treisman, 2010; Lidji et al., 2007; Rusconi et al., 2006). But the way in which pitch was mapped onto the pegboard in Mudd's experiment suggests that these two spatial dimensions are not mutually independent.

The finding that the mapping of tones in visual space was significantly associated with vertical and horizontal dimensions was sufficient for Mudd, who sought to develop coded auditory signals in order to "direct visual attention to the areas of an instrument panel that require immediate attention" (p. 347). But using the data reported in this study, it is possible to take the analysis a step further and examine the perception of pitch distances between tones, and how this relates to the spatial distances between pegs. If there is a common metric underlying representations of pitch space and visual space, then one would expect there to be a systematic relationship between perceived pitch distances separating reference and comparison tones and the spatial distances separating reference and comparison peg positions.

In order to do this it was first necessary to convert the hertz scale used to measure the frequency of tones to mel, which is a perceptual scale of pitches judged by listeners to be equal in distance from one another (Pedersen, 1965). As frequency measured in hertz increases, larger and larger frequency intervals are judged by listeners to produce equal pitch increments. It would therefore be problematic to compare frequency differences between tones with spatial differences between peg positions, as frequency does not have a linear relationship with listeners' perceptions of pitch. Mel conversions were calculated using the following formula taken from O'Shaughnessy (1987), where f equals frequency in hertz and m equals frequency in mel:

$$m = 2595 \log_{10}\left(1 + \frac{f}{700}\right)$$

Figure 3.2 demonstrates that there is a clear trend - the distance in mels between reference and comparison tones was positively correlated with the spatial distance (cm) between reference and comparison pegs, r = 0.79, n = 6, p = .03. However, visual inspection of the graph shows that the relationship may not be entirely linear. Interestingly, there also appears to be a different trend for the positioning of descending (1 to 3) and ascending (4 to 6) intervals. With descending intervals, the distance between peg positions increases exponentially compared with the increase in mel scale distance between tones. With ascending intervals, the opposite appears true – the distance between peg positions increases logarithmically compared with the increase in mel scale distance between tones.

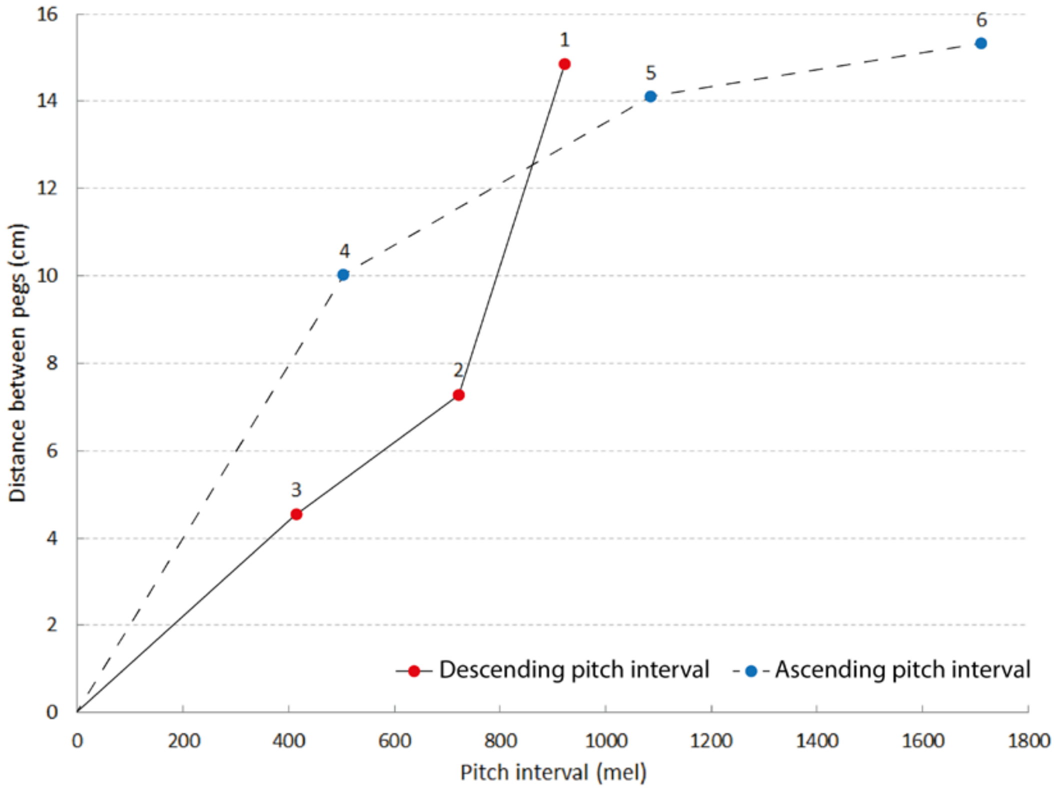


Figure 3.2. The relationship between perceived pitch difference (mel) visual distance (cm) reported by Mudd (1963). Numbered plots on the graph refer to the positions/pitches shown in Figure 3.1.

It is possible that this discrepancy is due to the methods employed in the experiments. Mudd presented participants with pure tones and, as stated elsewhere in this chapter, complex tones may be more suitable than pure tones when examining pitch perception. Also, the frequencies of tones used by Mudd were relatively high, so much so that the highest comparison tones were beyond the 5kHz upper threshold of efficient pitch perception (Attneave & Olson, 1971). It was reasoned that by taking into account these methodological issues, a replication of this experiment may allow the potential metric correlates of pitch space and visual space to be examined more thoroughly. Also, with advancements in technology, a study could be carried out in which responses were collected via a digital interface with much finer spatial resolution than a pegboard.

The aim of the current experiment was two-fold. Firstly, it served as a foundation for the experiments reported in Chapters 4, 5 and 6, seeking to demonstrate that the stimulus tones would be systematically plotted in visual space according to their perceived pitch. Secondly, it aimed to describe a common structural metric between pitch space and visual space. It was expected that the orientation of responses would be similar to those found in Mudd's experiment. The object representing comparison tones was expected to be repositioned on the vertical plane so that higher pitch comparison tones would be represented above the reference object, and lower pitch comparison tones would be represented below the reference object. To a lesser degree, the object representing comparison tones was also expected to be repositioned on the horizontal plane so that higher pitch comparison tones would be represented to the right of the reference object, and lower pitch comparison tones would be represented to the left of the reference object. In addition, it was expected that an increase in pitch interval between reference and comparison tones would be associated with an increase in the distance between reference and comparison objects. Of particular interest was the relationship between pitch height and distance between response plots. With the methodological changes made to Mudd's experiment it was expected that a more linear relationship would be observed between the scale of increase in mel and the scale of increase in cm.

3.3 Method

3.3.1 Design

A cross-modal psychophysical scaling task was employed in which the independent variable was the pitch of the comparison tone – and thus the interval size between the comparison and reference tones – and the dependent variable was the spatial positioning of the comparison object. The positioning of responses was measured in pixels along the x-axis and y-axis (pixels were converted to cm for the purposes of analysis). Ten tone pairs were presented in each block – there were 10 blocks altogether. The order of pairs was randomised within each block with the criterion that an interval change in the same direction was not presented any more than three times in succession.

3.3.2 Participants

An opportunity sample of 38 participants was recruited from the University of Roehampton by means of a course credit scheme. All participants were required to be at least 18 years of age, and to have normal or corrected-to-normal hearing acuity and vision. A short demographic questionnaire was completed by each participant, collecting information on gender, age, handedness, first language and musical training.

Participants were aged between 18 and 49 years (M = 24.26; SD = 7.68), 28 were female (74%), and all except 2 were right-handed. One participant reported a minor hearing problem (tinnitus) that was deemed not to have interfered with their ability to perform the task. The majority of participants (58%) reported English to be their first spoken language. Other first languages reported were Norwegian (5), Albanian (1), Chinese (1), Italian (1), Lithuanian

(1), Nepalese (1), Polish (1), Portuguese (1), Serbian (1), Slovak (1), Swedish (1) and Turkish (1). Finally, 53% of participants reported some level of musical training. Of these 30% reported more than 8 years, 10% reported 4 to 8 years, 25% reported 2 to 4 years, 5% reported 1 to 2 years, 10% reported 6 to 12 months and 20% reported less than 6 months.

3.3.3 Apparatus and stimuli

Stimuli were complex harmonic tones with a triangle waveform, synthesised in the same way as described for recognition experiments. The f_0 of tones were separated by intervals taken from the 5-note equal temperament tuning. 260.00 Hz was selected as the reference tone, and a further 10 frequencies in the range of an octave above and below this were selected as comparison tones: 1 = 130.00 Hz, 2 = 149.33 Hz, 3 = 171.54 Hz, 4 = 197.04 Hz, 5 = 226.34 Hz, 6 = 298.66 Hz, 7 = 343.07 Hz, 8 = 394.09 Hz, 9 = 452.69 Hz, 10 = 520.00 Hz. Ten .wav files of tone pairs were created in which the comparison tone was preceded by a reference tone. Each tone was 1000ms in duration with linear amplitude ramps applied to 50ms onset and 80ms offset. An ISI of 1500ms separated the two tones in every tone pair.

All visual stimuli were presented on a blank white screen with visual angle dimensions of approximately 35° width by 27° height. A solid black disc with a diameter of 1° was produced for the reference and comparison objects. A disc was chosen instead of a bar (as selected for visual stimuli in the recognition experiments) so as not to suggest a bias for the repositioning of the reference object in any particular direction. Participants were seated at a table in front of the computer screen. A chin rest was set up to ensure that their eye level was at the centre of the screen and at a distance of 40.00 cm.

All stimuli were generated and all responses were recorded on a Dell PC with Intel® Core ™2 CPU, 63000 @ 1.86GHz, 1.86 GHz, 1.99 GB of RAM, a Dell 1707 FP screen, and Shure SRH440 headphones at an intensity of approximately 60dB SPL. Panther (MIT) software was used to program the experimental procedure and data collection.

3.3.4 Procedure

Before starting the experimental trials (see Figure 3.3), participants were briefed on the task and gave informed consent. In each trial participants were presented with two consecutive pitches: a reference tone followed after a short pause by a comparison tone. At the same time as the reference tone was presented, the reference object appeared at the centre of the computer screen. A 50-ms linear ramp was applied to the onset of the reference object to match the onset ramp applied to the amplitude of the reference tone; rather than appearing immediately in full size on presentation, the reference object expanded to full size from a point at the centre of the screen over a time period of 50ms. Participants were told that the reference object represented the reference tone, and that their task was to indicate where on the screen the comparison tone should be represented, taking into account the position of the reference object. Participants could move the reference object from the centre of the screen by clicking on it with the mouse and dragging it. When they did this, a copy of the reference object remained at the centre so that when the task was finished, there were two objects on screen – a reference object that represented the reference tone, and a comparison object that represented the comparison tone. Once participants were happy with their response they pushed the spacebar to move on to the next trial, which started after a 1000ms pause.

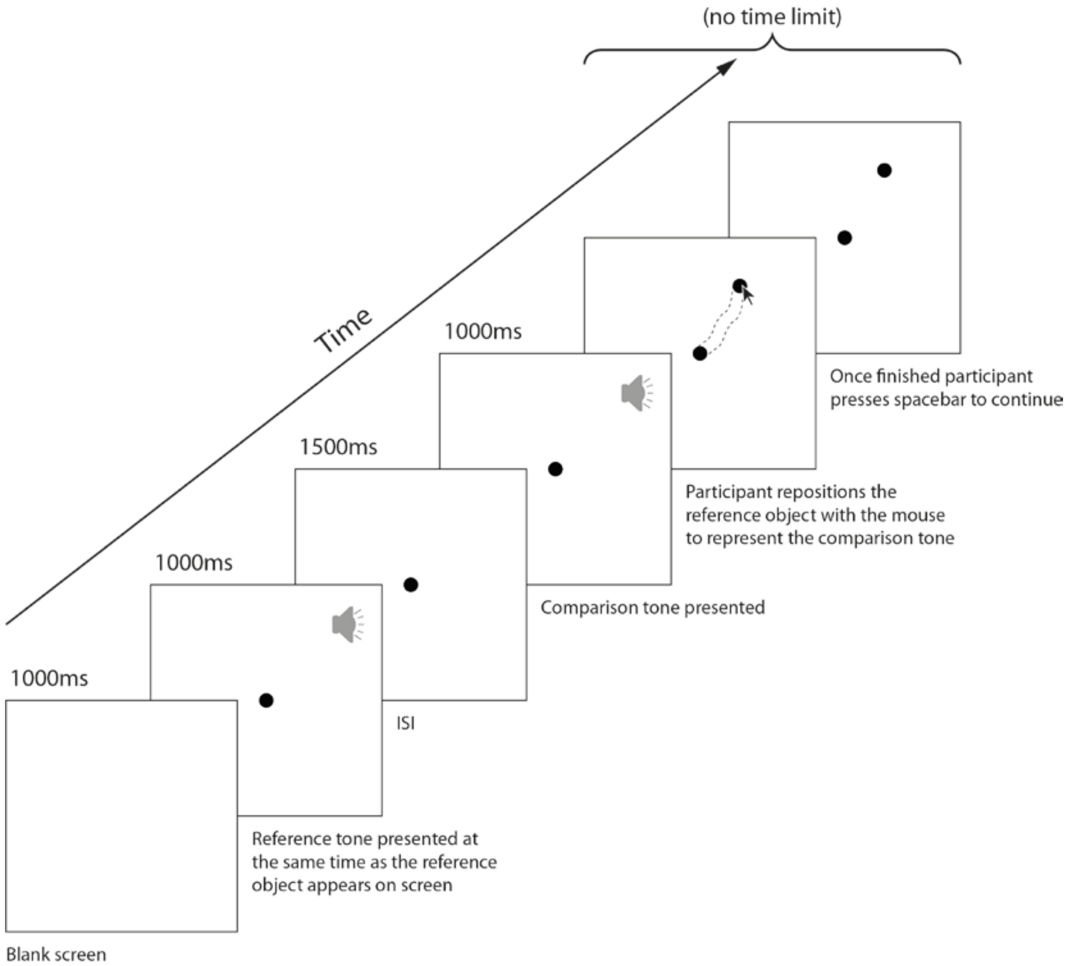


Figure 3.3. Experiment 1 trial timeline.

Care was taken to give minimal instruction to participants with regard to how to give their responses. Participants were simply told that they could move the object to any position on the screen, in any direction, and that they should try to be consistent throughout the trials. The experiment took approximately 20 minutes to complete.

3.4 Results and discussion

The first 10 trials were practice trials and were not included in the analysis. This meant that in total, 90 responses were collected from each participant – 9 responses for each tone pair (3420 in total). The data were then collapsed by tone pair across all participants. The average positioning of comparison objects was then calculated (see Figure 3.4). The trend of the results appears to be as expected – tones with higher pitch were represented as being higher in visual space, and further to the right compared with the reference tone. Similarly, tones with lower pitch were represented lower in visual space and further to the left. In other words - as pitch rises, spatial mappings move up the vertical axis and right along the horizontal axis; as pitch lowers, spatial mappings move down the vertical axis and left along the horizontal axis. These results replicate Mudd's (1963) original findings - the way in which pitch was mapped onto the computer screen was similar to how pitch was mapped onto the pegboard. When asked to place objects that represent tones of varying pitch in two-dimensional space, there is a tendency to systematically map pitch onto the vertical and horizontal axes. At the same time the vertical and horizontal spatial dimensions do not appear to be mutually independent, confirming the earlier conclusion that pitch is mapped onto a quasi-space that uses only two quadrants (the lower-left and the upper-right).

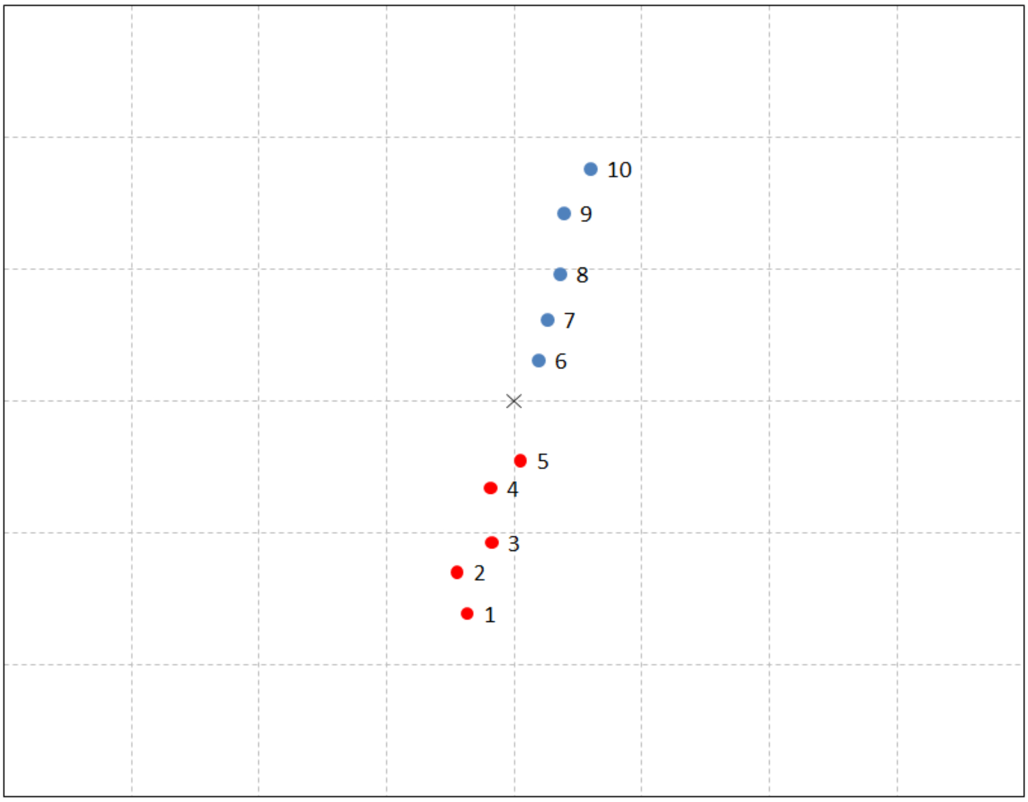


Figure 3.4. The averaged positioning of comparison objects by participants, representing different comparison tones (1 = 130 Hz, 2 = 149.33 Hz, 3 = 171.54 Hz, 4 = 197.04 Hz, 5 = 226.34 Hz, 6 = 298.66 Hz, 7 = 343.07 Hz, 8 = 394.09 Hz, 9 = 452.69 Hz, 10 = 520.00 Hz). The reference tone (260.00 Hz) was represented by a reference object (x) positioned at the centre of the screen (25.40 cm x 19.00 cm).

3.4.1 The relationship between pitch space and visual space

Next, the relationship between pitch distance and visual distance was examined. This relationship was first examined using the mel scale as a measure of perceived pitch interval. The analysis was subsequently repeated using the 5-note equal temperament scale as a measure of perceived pitch interval.

3.4.1.1 Pitch interval measured on the mel scale

Looking at Figure 3.5, there appears to be a clear linear relationship between pitch distance (mel) and visual distance (cm) - confirmed by a highly significant positive correlation, $r = 0.93$, $n = 10$, $p < .001$. A simple regression analysis was run to examine the extent to which interval size predicted distance between the positions of reference and comparison objects. Interval size explains a highly significant proportion of variance in the distance between objects, $F(1,378) = 207.40$, $p < .001$. The model explains 35% of the variance (Adjusted $R^2 = .35$). Interval size was a significant predictor of plot distance, $b = 0.02$, $t(379) = 14.40$, $p < .001$. The regression coefficient (b) for the interval size variable indicates that, as interval size increases by one unit, distance between the positioning of objects increases by 0.02 units. Therefore, an increase of 100 mel predicts a distance increase of 2.31 cm between objects (this is equal to a visual angle of 3.31° at the viewing distance of 40 cm).

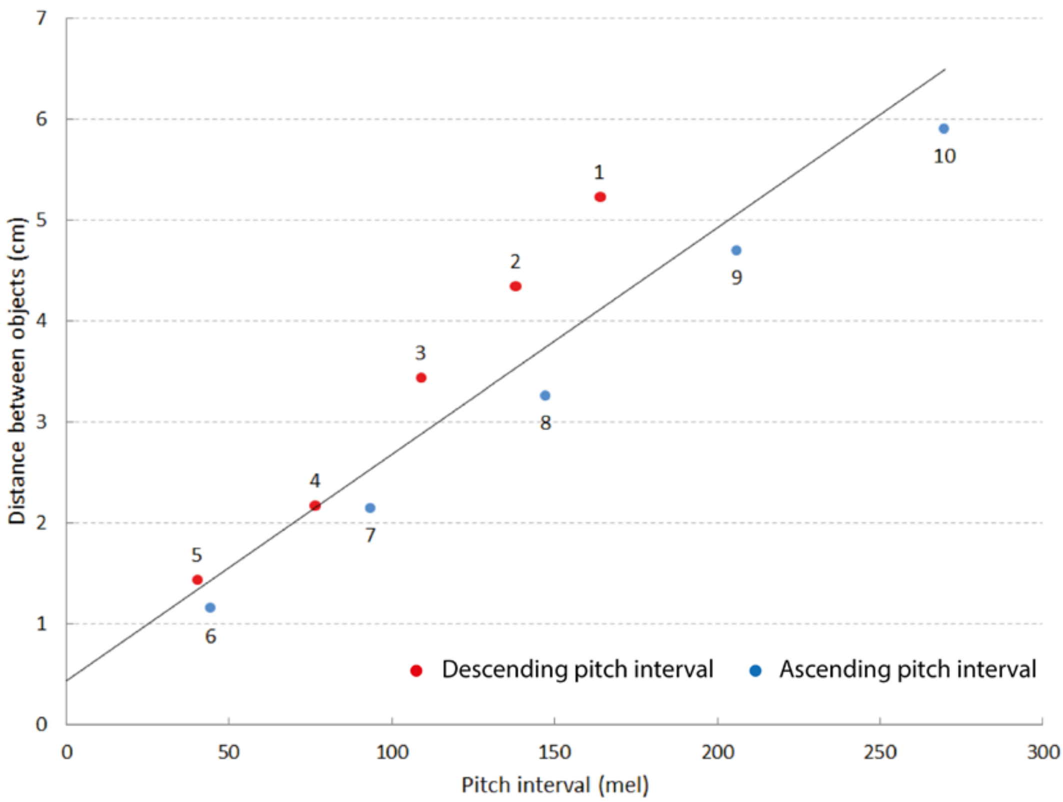


Figure 3.5. The relationship between the size of the visual distance separating reference and comparison objects, and the interval size separating reference and comparison tones (mel scale). Numbered plots refer to the digits assigned to comparison tones – see Figure 3.4.

It should be noted that the visual distances representing descending and ascending pitch intervals appear to have been mapped slightly differently – pitch distance for descending intervals was mapped onto a different slope compared with ascending intervals. For this reason a second multiple regression analysis was carried out which included interval direction (descending, ascending) as a categorical predictor variable, with ascending intervals as the reference group. Once again the model explained a highly significant proportion of variance in the distance between objects, $F(2,377) = 120.90$, $p < .001$. 39% of the variance was explained by the model (Adjusted $R^2 = .39$), slightly more than when not including interval direction as a predictor variable. As before, interval size was a highly significant predictor of plot distance, $b = 0.03$, $t(379) = 15.54$, $p < .001$. There was also a highly significant difference between predictions made by ascending and descending predictors – with descending intervals predicting greater distance between objects than ascending intervals, $b = 1.08$, $t(379) = 4.75$, $p < .001$.

An ANOVA was run to compare the first model with the second. It was found that the model including interval direction as a categorical predictor was highly significantly better than the model that did not, $F(1,377) = 22.59$, $p < .001$. From this it can be concluded that, in terms of comparing spatial distance with pitch distance measured on the mel scale, there appears to be a difference in the way descending and ascending pitch intervals are mapped onto visual space.

The mean distances in response to intervals in different directions were not very different – mean distance for descending intervals was 4.18 cm and mean distance for ascending intervals was 4.30 cm. Why then, did the above regression model predict a difference in distances between descending and ascending intervals of 1.08 cm? The answer can be found in the mean interval sizes. For descending intervals the mean size was 105.53 mel, and for ascending intervals the mean size was 152.33 mel. Thus, measured in mel scale, descending intervals were on average smaller than ascending intervals, but mean distances for intervals in both directions were similar.

3.4.1.2 Pitch interval measured on the 5-note equal temperament scale

Attneave and Olson (1971) proposed that the mel scale is not the best measure of pitch scaling, as it does not reflect the invariances preserved in transposition. In Figure 3.6, the data were re-plotted so that pitch was measured in terms of the 5-note equal temperament scale instead of mel scale. This time the residuals were smaller and although it appears that a slight difference persists, descending and ascending intervals are seemingly mapped onto a very similar slope. In other words, there no longer appears to be a difference in the way descending and ascending pitch intervals are mapped onto visual space.

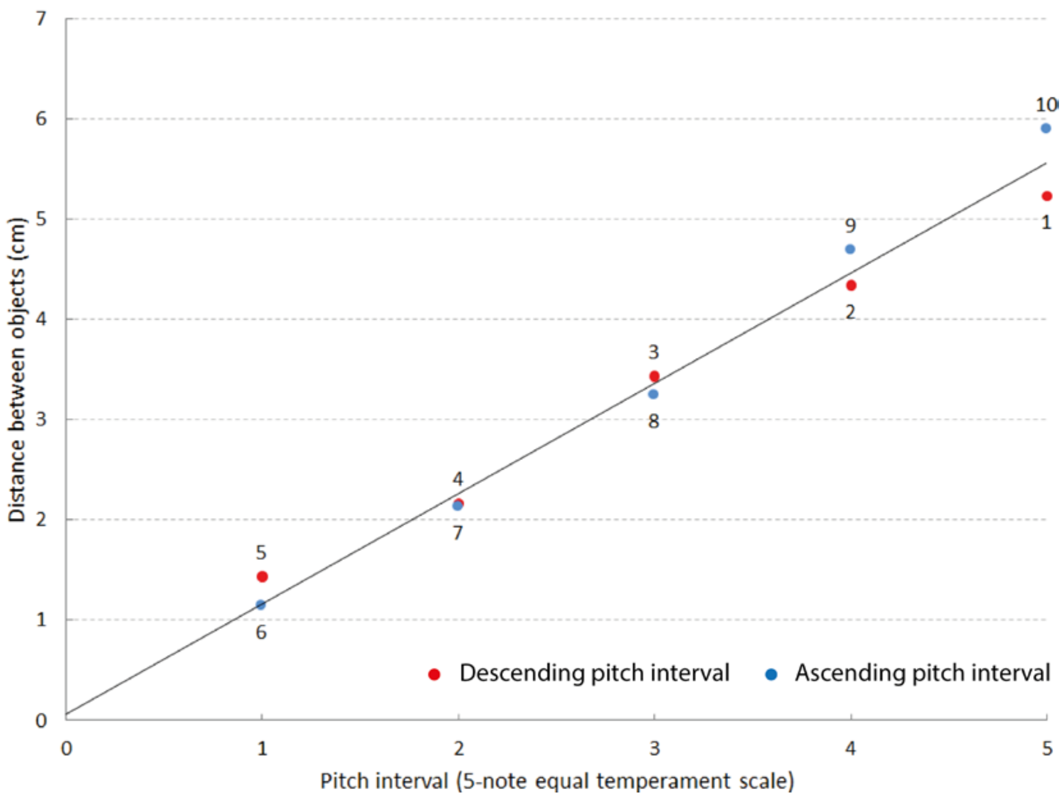


Figure 3.6. The relationship between the size of the visual distance separating reference and comparison objects, and the interval size separating reference and comparison tones (5-note equal temperament scale). Numbered plots refer to the digits assigned to comparison tones – see Figure 3.4.

As before, there was a highly significant positive correlation between the two variables, $r = 0.99$, $n = 10$, $p < .001$. First, a simple regression analysis was run to examine the extent to which interval size predicted distance between the positions of reference and comparison objects. Interval size explains a highly significant proportion of variance in the distance between objects, $F(1,378) = 262.81$, $p < .001$. The model explains 41% of the variance (Adjusted $R^2 = .41$). Interval size was a highly significant predictor of plot distance, $b = 1.21$, $t(379) = 16.21$, $p < .001$. The regression coefficient (b) for the interval size variable indicates that as interval size increases by one unit, distance between the positioning of objects increases by 1.21 units. Therefore, an increase in one interval of the 5-note equal temperament scale predicts a distance increase of 1.21 cm between objects (this is equal to a visual angle of 1.35° at the viewing distance of 40 cm).

Importantly, Figure 3.6 suggests that both descending and ascending pitch intervals of the 5-note equal temperament scale were mapped onto visual space relatively uniformly. In order to test this, a second multiple regression analysis was carried out which included interval direction as a categorical predictor variable. Once again the model explains a highly significant proportion of variance in the distance between objects, $F(2,377) = 131.40$, $p < .001$. However, the model explains the same amount of variance (41%) as the previous model that did not include interval direction as a predictor variable (Adjusted $R^2 = .41$). Interval size was a highly significant predictor of plot distance, $b = 1.21$, $t(379) = 16.20$, $p < .001$. Crucially, the difference between predictions made by ascending and descending predictors was not significant, $b = -0.13$, $t(379) = -0.60$, $p = .55$. An ANOVA confirmed that the model including interval direction as a categorical predictor was no better than the model that did not, $F(1,377) = 0.36$, $p = .55$.

This result may be explained by considering the stimuli used in this experiment. Firstly, the highest and lowest frequencies of presented comparison tones (tones 1 and 10) were octaves of the reference tone. An octave is the most recognisable of pitch intervals and it is reasonable to assume that all participants recognised these as the extreme values of the variable each time they heard them. The remaining comparison tones could then be represented by objects positioned in order of their perceived pitch height in between these 'book ends' with equal spacing. In other words, octaves served as anchors and the space between them was partitioned linearly. Although the 5-note equal temperament scale was adopted as a way of trying to avoid the effects of tonality on perceptions of pitch space, it appears that the tonal references provided by the octave intervals were sufficient for the participants to 'tune in' to the tonal relationships of this alternative chromatic tuning.

3.4.2 Addressing the diagonal shift

In order to examine the issue of the diagonal shift observed in the averaged positioning of comparison objects, a closer inspection of results for individual participants was carried out. Although there were many participants who demonstrated the same diagonal mapping of pitch onto visual space as observed in the overall findings, there were groups of participants who adopted sometimes very different approaches, as can be seen in Figure 3.7 (see Appendix II for all individual participants' averaged mappings). The greatest proportion of participants mapped pitch onto a purely vertical representation (Figure 3.7A). Others mapped pitch onto a diagonal representation (Figure 3.7B). Others still adopted completely unique approaches, mapping higher pitch to the left of the reference plot (Figure 3.7C) or even below it (Figure 3.7D). In short, there was a large degree of variability in the way in which participants mapped pitch space onto visual space which is obscured by the averaging of results. Although the diagonal shift is evidenced in the responses of some individual participants, the overall trend must be attributed to a number of different spatial response patterns. These include the dominant vertical dimension as well as a horizontal dimension, but they also include other, sometimes unique, mappings.

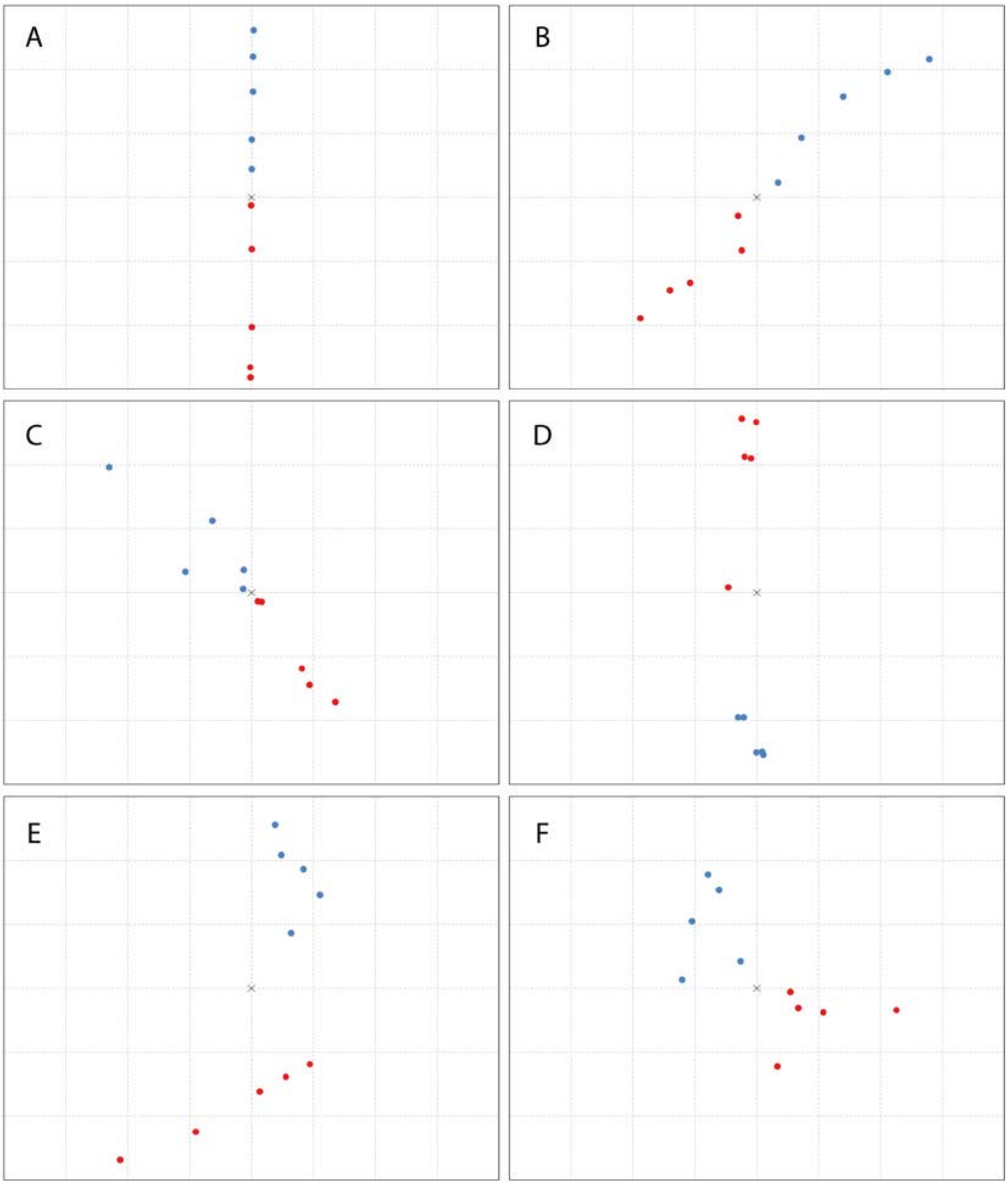


Figure 3.7. The averaged positioning of comparison objects by six different participants. Red dots indicate responses to descending pitch intervals and blue dots indicate responses to ascending pitch intervals. (A) and (B) demonstrate typical responses on the vertical and diagonal axes, with mapping on a linear scale. (C) and (D) demonstrate atypical responses on alternative diagonal and inverted vertical axes. (E) and (F) demonstrate inconsistent/disordered mappings.

Another observation from Figure 3.7 is that pitch was not always mapped onto a linear trajectory. It appears that a number of participants mapped pitch onto visual representations that were dispersed across the screen (Figure 3.7E and F). While it is not immediately clear why such mappings were made, comments collected from participants indicated that the mapping of tones was not always made exclusively according to their perceived pitch height. Some participants reported that they felt some tones were more dissonant than others. In these instances it is plausible that participants were trying to relate their knowledge of tonality to the stimuli. As the 5-note equal temperament scale was used, tonal relationships were avoided to some extent. However, some intervals are unavoidably closer to those of a diatonic scale than others. As a result, some participants may have been mapping pitch according to more than one parameter – for example, pitch height and tonal consonance.

Inspection of individual response data has demonstrated that participants exhibited a number of different mapping strategies. Irrespective of this, it remains possible that participants were more consistent in mapping pitch interval size onto visual distance, than they were in mapping pitch interval direction onto a particular trajectory in visual space. In order to examine this possibility, additional analyses were carried out comparing variance in the trajectory of comparison plots with variance in the distance between comparison plots. Variance in trajectory was calculated by measuring the angle between adjacent comparison objects (i.e. those that represent adjacent tones on the pitch scale) in relation to the reference object (see Figure 3.8A). Variance in distance was calculated by measuring the absolute distances between adjacent comparison objects (see Figure 3.8B). The resulting angles and distances were then standardised by converting to ratios, and their variance was calculated.

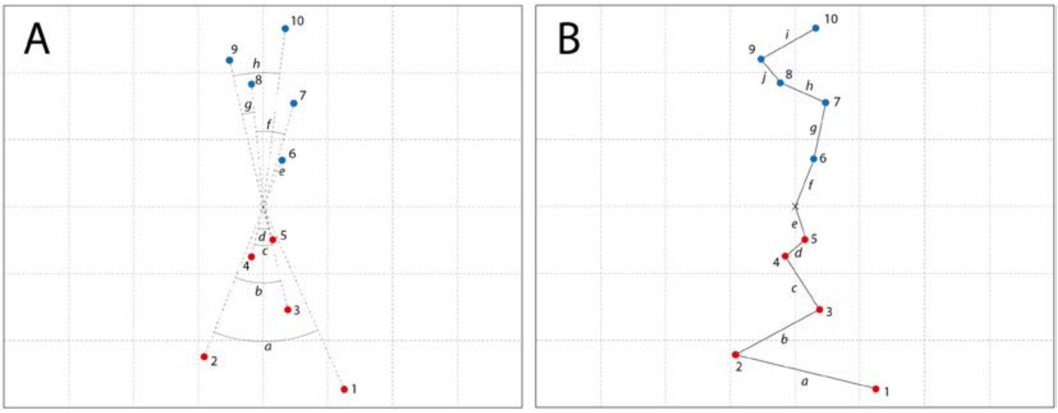


Figure 3.8. Illustration of the measurements applied to compare variability in the direction of mappings (A) with the distance between mappings (B) of an individual participant's responses.

The variance in angle between comparison objects (M = 1.92*10^{-2}, SD = 1.68*10^{-2}) was significantly greater than the variance in distance between comparison objects (M = 0.41*10^{-2}, SD = 0.33*10^{-2}), $t(37) = 5.80$, $p < .001$, two-tailed. This suggests that when participants were mapping pitch onto visual space, the relationship between pitch interval and the distance between mapped objects was more stable than the relationship between pitch and the direction of mapped objects; i.e. participants were more consistent in their mappings of pitch distance onto visual distance than they were with their mappings of pitch height onto a visual trajectory. An important implication of this finding for the present research is that, when constructing analogous pitch and visual stimuli, representing pitch height on a particular visual axis may be less important than representing pitch interval size on a particular visual scale. This interpretation confirmed the decision to represent pitch height on the vertical dimension, as this was the strongest representation observed across individuals.

3.5 Conclusion

The present experiment investigated the potential existence of a common metric for pitch space and visual space. Participants took part in a cross-modal scaling task in which they were presented with a reference tone that was represented by a reference object at the centre of a computer screen. They were then presented with a comparison tone and instructed to position a comparison object on screen to represent this second tone. The results confirm previous findings and indicate that pitch is systematically mapped onto a vertical dimension, and also to a lesser extent onto a horizontal dimension. In addition, the size of pitch intervals separating presented tones appears to have a linear relationship with the size of the distance separating the positions of corresponding objects in visual space. The relationship between pitch space and visual space was best explained when pitch distance judgments were measured in terms of the 5-note equal temperament scale relative to the mel scale. Furthermore, the mapping of pitch interval size onto visual distance was more stable than the mapping of pitch interval direction (ascending, descending) onto a visual trajectory. Further analysis demonstrated that an increase of one interval of the 5-note equal temperament scale predicted an increase in visual distance between reference and comparison objects of 1.21 cm, which translates to a visual angle of 1.35˚ from the viewing distance of 40 cm. As a result of these findings, all subsequent experiments used analogous auditory and visual stimuli in which pitch interval corresponded to visual angle at the ratio of 1:1.35.

Chapter 4: Transformation recognition in non-equivalent supramodal pattern spaces

4.1 Introduction

Experiments 2 and 3 are reported in the present chapter. The general aim of these experiments was to explore the possibility that shared cognitive mechanisms are involved in the mental transformation of sequential pattern structure within the theoretical framework outlined in Chapter 2. In particular, the investigation focussed on the processing of structural information abstracted from auditory pitch sequences and analogous visuo-spatial sequences that had undergone one of two types of isomorphic transformation: inverse and retrograde. As a starting point, the experiments investigated the perception of transformations when auditory and visual stimuli corresponded to different supramodal pattern spaces.

The motivation for the aim of the present experiments is based on a number of different areas of psychological research that point towards the possibility that a supramodal mechanism (or mechanisms) is involved in the processing of structural transformations. Classic research in serial pattern learning made extensive use of rules that described transformations equivalent to the inverse and retrograde transformations under examination in the present experiments (inverse transformation: mirror image [M] or complement [C] rule; retrograde transformation: inversion [I] rule), and numerous previous experiments have demonstrated that participants are able to use these rules to produce hierarchically organised representations of patterns (e.g. Jones, 1976a; Kotovsky & Simon, 1973; Leeuwenberg, 1968; Restle & Brown, 1970; Restle, 1970, 1973, 1976; Simon & Kotovsky, 1963). Theories of serial pattern learning were not specific to auditory or visual sensory modalities, and the assumption of these theories was that the learning of pattern structure was a general process.

Inverse and retrograde transformations have also been applied to melodic structure in music composition and for this reason have received some interest from music psychologists (e.g. Dienes & Longuet-Higgins, 2004; Dowling, 1971, 1972; Krumhansl, Sandell, & Sergeant, 1987; Schulze, Dowling, & Tillmann, 2012; White, 1960). Recent behavioural findings in this field of research have suggested that the processing of melodies under these structural transformations may be associated with spatial processing (Cupchik, Phillips, & Hill, 2001; McLachlan, Greco, Toner, & Wilson, 2010). For instance, recognition performance in a melodic transformation task has been correlated with performance in a visuo-spatial mental rotation task (Cupchik et al., 2001). In support of these behavioural data, recent brain imaging data have demonstrated that shared higher-order areas in the cortex (specifically, the intraparietal sulcus [IPS]) are activated by the mental transformation of melody and the mental manipulation of visuo-spatial information (Foster, Halpern, & Zatorre, 2013; Zatorre, Halpern, & Bouffard, 2010). All of this evidence is converging on the possibility that inverse and retrograde transformations of melody may be processed by shared mechanisms responsible for the processing of spatially represented structural information.

Though previous researchers, such as those cited above, have proposed the possibility of partly shared representations and processing of auditory pitch patterns and visuo-spatial patterns, to date a thorough and systematic investigation has not been carried out. In order to address this issue, a theoretical framework has been proposed that conceives of a supramodal pattern space (SPS; see Chapter 2, Section 2.2). Representations of patterns in this space are determined by structural information that has been abstracted from auditory and visual sensory information. Depending on the nature of the stimulus, representations in supramodal pattern space are constructed on one or a combination of two types of qualitatively distinct supramodal dimensions: a scalar and a temporal dimension. The scalar dimension corresponds to the relative pitch of tones, or the relative position of visual objects. The temporal dimension corresponds to the relative timing of tones or the relative timing of objects in a sequence. Importantly, whilst the scalar dimension is conceived to be fully spatial, in that pattern events can freely move in both directions along this dimension, the temporal dimension is held to retain some level of directionality, due to the inherent directionality of temporal relations.

The SPS framework provided a means of describing equivalent (or non-equivalent) representations of auditory and visual stimuli. Simple auditory pitch patterns, such as the ones used in the present research, can be represented in a 1½-D supramodal pattern space, constructed from a scalar dimension (which represents relative pitch) and a temporal dimension (which represents the relative timing of tones). One of the assumptions of the SPS framework is that any process contributing to the perception of structural regularities described by isomorphic transformation requires an inversion of ordinal relations on a supramodal dimension. Inversions on a scalar dimension are proposed to be relatively easy to process, due to the dimension's bidirectionality – an inversion of ordinal relations would not be incompatible with the dimension's directionality. Inversions on a temporal dimension, on the other hand, are proposed to be relatively difficult to process. Inversions of ordinal relations on this dimension are incompatible with the dimension's unidirectionality, which would imply some resistance to the transformation.

For patterns represented in a 1½-D supramodal pattern space, inverse transformations require an inversion of ordinal relations on a scalar dimension, and retrograde transformations require an inversion of ordinal relations on a temporal dimension. Thus, the 1½-D hypothesis predicted that, when based on the processing of structural information that can be represented in a 1½-D supramodal space, pattern regularities described by inverse transformation should be perceived more effectively than pattern regularities described by retrograde transformation.

For representations of visual stimuli to also correspond to a 1½-D supramodal pattern space, they would need to consist of a sequence of objects presented at different locations on a single spatial dimension. However, the majority of previous research investigating the processing of auditory pitch patterns and structurally analogous visual patterns has mapped the pitch of auditory tones onto the vertical axis in visual space, and their timing onto the horizontal axis (Balch & Muscatelli, 1986; Billig & Müllensiefen, 2012; McLachlan et al., 2010; Prince, Schmuckler, & Thompson, 2009). This has been done even when visual stimuli are presented sequentially (Balch & Muscatelli, 1986).

When presenting an analogous visual pattern as a single two-dimensional image, it makes sense to map time onto the horizontal axis – as has already been reported in this thesis, there is a growing body of experimental evidence supporting the mental representation of time in horizontal space (Ishihara, Keller, Rossetti, & Prinz, 2008; Lakens, Semin, & Garrido, 2011). But, when pattern events are presented sequentially the dimension of time is already inherent in the stimulus. By presenting events at different positions on the vertical and horizontal axes, this in effect introduces an additional spatial dimension. This much becomes clear when thinking about how structural information, abstracted from a horizontally presented sequential visual stimulus, would be represented in a supramodal pattern space. A scalar dimension would represent the vertical height of objects, a temporal dimension would represent the timing of visual events, and an additional scalar dimension would be needed to represent their horizontal position. Therefore, structural information would be represented in a 'two-and-a-half-dimensional' (2½-D) supramodal pattern space (see Figure 4.1), as opposed to a 1½-D supramodal pattern space.

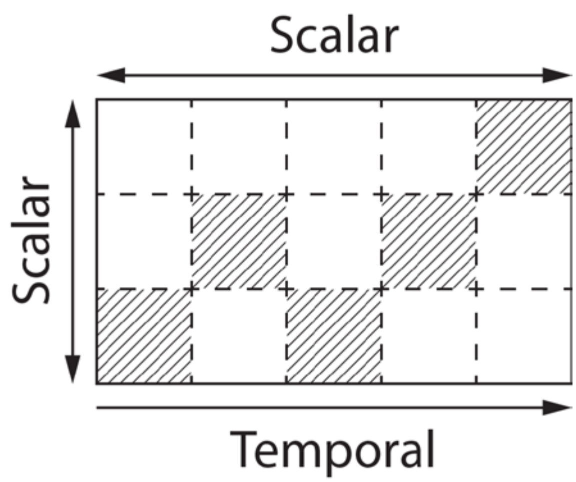


Figure 4.1. 2½-D supramodal pattern space.

This has important implications for mental transformation. It has been proposed that retrograde transformations of patterns represented in a 1½-D space will be harder to process because they require an inversion of ordinal relations on a directional temporal dimension. Retrograde transformations of patterns represented in a 2½-D space, on the other hand, could require an inversion of ordinal relations on a directional temporal dimension, but also an inversion of ordinal relations on a bidirectional scalar dimension. As such, they would be no more difficult to process than inverse transformations, which also require an inversion of ordinal relations on a bidirectional scalar dimension. Furthermore, they may even be processed more efficiently, because the structural relations that need to be transformed are represented on not one, but two supramodal dimensions. This introduces additional redundancy which has been linked to better perception (Hochberg & McAlister, 1953).

So, when analysed within the SPS framework, it is clear that the way in which visual stimuli are presented is a key issue when investigating possible shared representations and processing of auditory pitch patterns and visual patterns. Visual stimuli that map time onto the horizontal dimension may be represented differently to visual stimuli that do not map time onto the horizontal dimension, and present all sequential components at the same central location. In turn, these representations may be processed differently. However, to date this issue has not been properly addressed, and this is reflected in the fact that different researchers have presented visual stimuli in different ways.

To summarise the above discussion, it is proposed that auditory pitch patterns correspond to a 1½-D supramodal pattern space. However, previous research that has made a structural analogy between auditory and visual stimuli has presented visual patterns that correspond to a 2½-D supramodal pattern space. Two contrasting predictions have been made for the perception of structural transformations in different supramodal spaces. The 1½-D hypothesis predicts a processing advantage for inverse transformations. The 2½-D hypothesis, on the other hand, predicts no processing advantage for inverse transformations, and a potential processing advantage for retrograde transformations.

The experiments reported in the present chapter represent an initial attempt to examine potential supramodal processes using the SPS framework. As a starting point, they sought to investigate the processing of auditory and visual patterns when they corresponded to different pattern spaces – i.e. when auditory patterns corresponded to a 1½-D and visual patterns corresponded to a 2½-D supramodal pattern space. In order to test the 1½-D and 2½-D hypotheses, the experiments employed a short-term recognition paradigm (see Chapter 2, Section 2.4 for rationale). The paradigm involves the presentation of a standard pattern followed after a short pause by a target pattern. In the experiments reported here, the target pattern was either related to the standard under inverse or retrograde transformation or it was unrelated under these types of transformation. The task was to indicate whether the target was related or unrelated to the standard. Theoretically, a target could only be recognised successfully by perceiving its relationship (or lack of relationship) to the standard by processing the relevant structural transformation. Therefore performance in the task reflected participants' ability to perceive pattern relationships between the standard and target patterns.[16] In order to examine performance, accuracy and response times were analysed. Lower error rates and faster response times were assumed to indicate more efficient perceptual processing.

[16] This, of course, did not exclude the possibility that performance also reflected memorial failures rather than perceptual ones.

4.2 Experiment 2: Unimodal trials

The aim of Experiment 2 was to examine the hypotheses outlined above, in a unimodal context (i.e. standard and target patterns were presented in the same sensory modality). In each trial, the standard and target patterns were either both auditory (A), or they were both visual (V).

Due to the way in which stimuli were presented, different hypotheses were tested in both modality conditions. According the SPS framework, auditory pitch patterns can be represented in a 1½-D supramodal pattern space. The 1½-D hypothesis predicted that, when target recognition is based only on the processing of structural information, then recognition performance should be better for inverse transformations. In keeping with previous research, structurally analogous visual patterns consisted of objects presented sequentially at different locations in two-dimensional visual space. The pitch of tones was mapped onto the vertical dimension and the timing of tones was mapped onto the horizontal dimension. Visual stimuli presented in this way can be represented in a 2½-D supramodal pattern space. Therefore, in the visual condition the 2½-D hypothesis was tested. This predicted that recognition performance for inverse transformations should be no better than for retrograde transformations, and that recognition performance for retrograde transformations may be better than for inverse transformations.

4.2.1 Methods

4.2.1.1 Participants

54 students from the University of Roehampton took part in Experiment 2 (female = 42, male = 12; mean age = 21.28 years, *SD* = 5.67). All had normal hearing and normal or corrected-to-normal vision. Four participants reported they

were left-handed, one reported they were ambidextrous and the remainder were right-handed. 20 participants (37%) reported some previous music training (mean = 3.48 years). They all received course credit for their participation.

4.2.1.2 Stimuli

In each experimental trial, participants were presented with either a pair of auditory pitch patterns or a pair of visual patterns. The generation of pattern structures used in the experiment is detailed in Chapter 2 (Section 2.5.1) of this thesis (see Figure 4.2 for an example pattern structure and its corresponding auditory and visual realisations). Briefly, the patterns were sequences of 5 events and were ternary (3 possible values). An auditory pitch pattern was a sequence of 5 tones that were one of 3 different pitches. A visual pattern was a sequence of 5 black bar segments that were presented at one of 3 different vertical heights. In each case, the first pattern (henceforth "standard") was either structurally related or unrelated to the second pattern (henceforth "target"). Targets that were related to the standard possessed the same structure that had undergone either a retrograde or an inverse transformation. Targets that were unrelated possessed different structure and were therefore not related under inverse, retrograde, or retrograde inverse transformation.

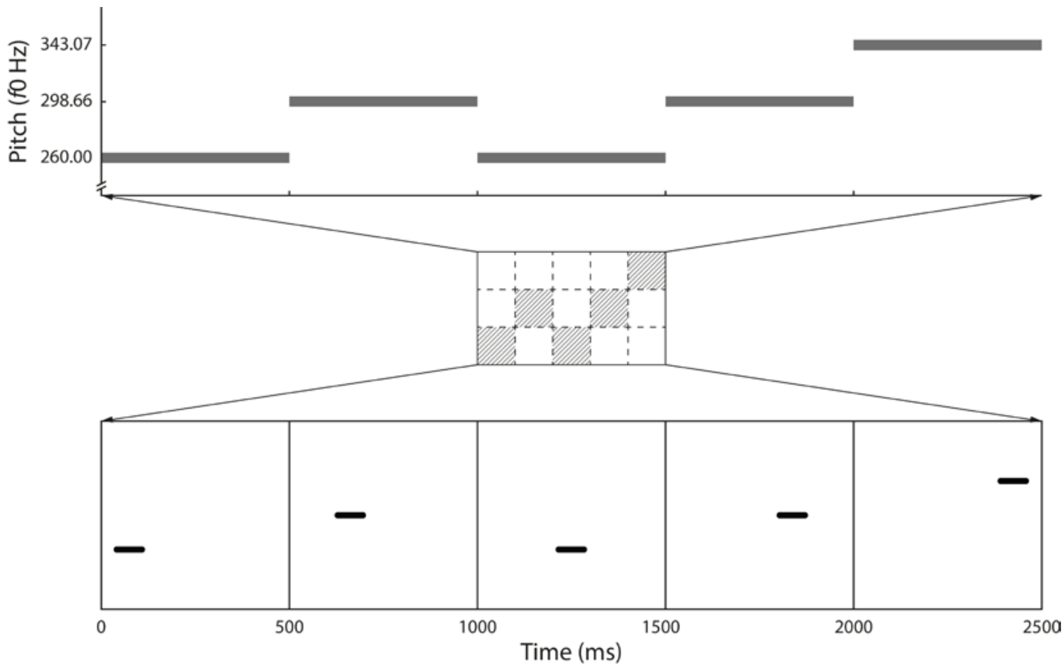


Figure 4.2. Auditory and visual stimuli used in Experiment 2. Top: A time-frequency plot of an auditory stimulus. An auditory stimulus consisted of a sequence of 5 tones presented at one of 3 different pitch heights. Bottom: Each panel displays a still image that was presented for 500ms. A visual stimulus consisted of a sequence of black bar segments presented at one of 3 different vertical heights. Segments changed positions from left to right along the horizontal axis.

Auditory pitch patterns were monophonic, isochronous and composed from a 5-note chromatic scale. All standard and target melodies shared the same three pitches with fundamental frequencies of 260.00 Hz, 298.66 Hz and 343.07 Hz. Tones were produced to have complex triangular waveforms. Each tone was 500ms in duration with onset and offset ramps of 10ms. The onset of each successive tone of a sequence occurred on the offset of the preceding tone. A complete sequence was 2500ms in duration. Auditory stimuli were presented binaurally through Shure SRH440 headphones at a comfortable listening level of approximately 60 dB SPL.

Visual patterns were designed to be analogous to pitch patterns – tones of different pitch were replaced with black bars presented at different vertical heights on a white background. Black bars measured 2.23˚ by 0.45˚ in size. They were presented from left to right in their temporal order with no gap between each successive bar (if two successive bars were to be presented simultaneously at the same height, the right end of the earlier bar would meet the left end of the later bar). The vertical distance between bars at different heights was 1.35˚ (as determined by Experiment 1). Thus, the maximum area covered by an entire visual sequence, from left to right, top to bottom, measured 11.15° by 3.15°. The middle temporal event (the third event of the sequence) was aligned to the centre of the display screen's horizontal axis. The middle height value was aligned to the centre of the display screen's vertical axis. Each bar was presented for 500ms in duration. The onset of each successive bar of a sequence occurred on the offset of the preceding bar. A complete sequence was 2500ms in duration.

4.2.1.3 Design and procedure

The design was within-subjects, with three asymmetrically arranged independent variables – two levels of modality (auditory, visual), two levels of relatedness (related, unrelated) and two levels of transformation (retrograde, inverse) embedded into the former related level. Consequently, there were six experimental conditions, and the proportion of trials per condition was as follows: 1) auditory, related, retrograde (ARR) = 12.5%; 2) auditory, related, inverse (ARI) = 12.5%; 3) auditory, unrelated (AU) = 25%; 4) visual, related, retrograde (VRR) = 12.5%; 5) visual, related, inverse (VRI) = 12.5%; 6) visual, unrelated (VU) = 25%.

15 standard patterns were used in the experiment (see Chapter 2, Section 2.5.1, Table 2.4). Each pattern was presented once in related conditions and twice in unrelated conditions, making a total of 120 trials per experimental session. The experimental session was divided into 60-trial blocks containing equal proportions of auditory and visual trials and equal proportions of related and unrelated trials. Participants only had to recognise one type of transformation per block, therefore one block contained all the related inverse trials, and the other block contained all related retrograde trials. Each block was further sub-divided by modality into 30-trial sub-blocks, containing 15 related trials and 15 unrelated trials each. The order of blocks was counterbalanced between participants. The order of sub-blocks within each block was randomised, as was the order of trials within each sub-block.

On arrival participants completed a brief questionnaire collecting demographic information pertaining to age, gender, handedness, potential hearing problems and musical experience. Participants were then seated in front of a PC monitor and taken through a series of on-screen instructions by the experimenter. Before each sub-block, participants were introduced to the relevant transformation and presented with examples (example stimuli were bimodal – i.e. they were presented simultaneously as auditory and visual patterns). Before each sub-block, they took part in 6 untimed example trials with the experimenter present. These trials were presented in the modality condition corresponding to the sub-block. They were instructed to focus on both patterns and decide whether the second pattern was a transformation of the first (‘related’) or ‘unrelated’ to the first. They indicated their decision by pressing one of two buttons on a response box using the index and middle fingers of their dominant hand. In addition, they were instructed to respond as quickly as possible whilst maintaining accuracy. In half of the experimental sessions 'related' responses were allocated to the left button, and in the other half 'related' responses were allocated to the right button. When the participant was ready to start the experimental trials the experimenter left the room.

Figure 4.3 illustrates the time course of a trial in the auditory condition. Trials in the visual condition were exactly the same, only auditory pitch patterns were replaced with visual patterns. At the beginning of each trial a fixation cross was displayed at the centre of the screen for 1000ms, and a 200ms warning beep (pure tone, 298.66 Hz) alerted the participant to the fact that the standard would be presented soon. Once the standard pattern had been presented, the fixation cross reappeared for a 2000ms ISI, before the target pattern was presented. On the offset of the target, participants indicated whether they thought the target pattern was related or unrelated to the standard. The next trial was initiated by the participants' response. If the participant had not responded after 5000ms the next trial started automatically. Before starting the experimental trials, participants took part in 6 timed practice trials. Feedback was provided for responses to practice trials, but not for responses to the experimental trials. Between each block and sub-block the experimenter returned to provide further instructions specific to the upcoming trials.

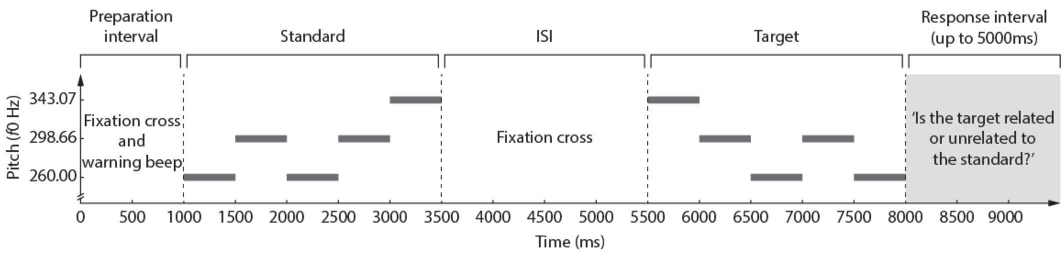


Figure 4.3. Time course of a trial in the auditory condition, Experiment 2. The target is a retrograde transformation of the standard. Targets could also be inverse transformations or unrelated to the standard.

The experiment was carried out in a sound-attenuated room. Background noise level was minimal, not exceeding 30 dB SPL, and ambient luminance was kept at a constant low level in order to provide a non-distracting experimental environment. At the end of the experimental session, all participants were debriefed and encouraged to comment on their experience.

4.2.2 Results

Data from 1 participant were excluded from analysis because they failed to perform above chance levels on the task (their overall error rate was 50% or greater). Prior to the main analysis, paired-samples *t*-tests were run on block order (first block, second block) to examine any effects of learning on overall PE (arcsine-transformed) and RT (log-transformed). The order of blocks was counterbalanced between participants by transformation level, but it was possible that due to the novelty and difficulty of the task, a general trend would be observed whereby participants improved significantly after having practised the transformation task in the first block. Both *t*-tests failed to reveal a significant effect of block order (PE: $t(52) = 0.78$, $p = .219$, one-tailed; RT: $t(52) = 0.42$, $p = .340$, one-tailed).[17]

4.2.2.1 Error data

4.2.2.1.1 All trials

Overall percentage error (PE) was 30.49. This was consistent with previously published behavioural data using a similar task in a melody recognition experiment (Dowling, 1972). Mean results for responses to inverse, retrograde and unrelated targets in both modality conditions are displayed in Figure 4.4. In the first stage of the analysis an ANOVA was run on arcsine-transformed PE data to examine the within-subjects effects of relatedness (related, unrelated) and modality (auditory, visual). This was done in order to ascertain whether similar processing strategies were employed in responding to related and unrelated stimuli. If participants responded to unrelated targets in the same way as they responded to related targets, there would be little purpose in attempting to establish a difference between transformation conditions. The absence of a significant difference between related and unrelated conditions would have implied that the design of the experiment was inappropriate.

[17] NB Mean data and ANOVA tables for all experiments can be found in Appendix III.

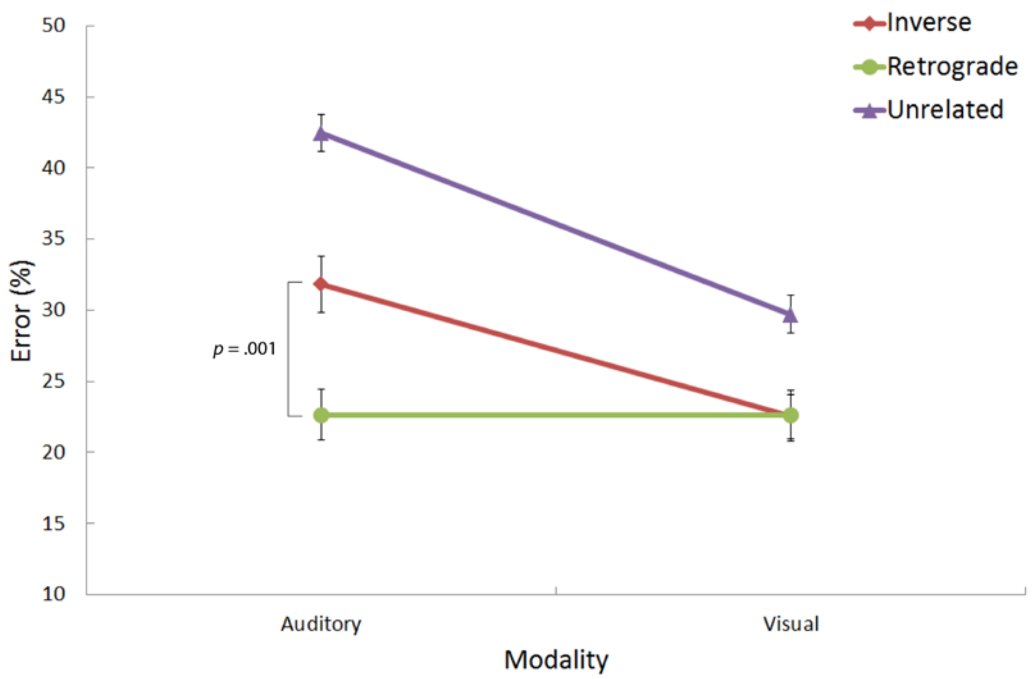


Figure 4.4. Experiment 2: Mean PE in target conditions, plotted as a function of modality. Significance values for simple effects of transformation were obtained from the analysis on arcsine-transformed data (NB simple effects of modality are not displayed). Error bars indicate the standard error of the mean (SEM ± 1).

The 2 x 2 ANOVA revealed a highly significant main effect of relatedness, $F(1,52) = 38.34$, $MSE = .02$, $p < .001$, $\eta_p^2 = .42$. Participants made more errors when responding to unrelated targets ($M = 36.09$, $SD = 12.79$) compared to related targets ($M = 24.90$, $SD = 10.45$). The main effect of modality was also highly significant, $F(1,52) = 33.31$, $MSE = .02$, $p < .001$, $\eta_p^2 = .39$, with participants making more errors in the auditory ($M = 34.84$, $SD = 10.78$) compared to the visual condition ($M = 26.14$, $SD = 11.52$). There was a significant interaction between relatedness and modality, $F(1,52) = 9.76$, $MSE = .01$, $p = .003$, $\eta_p^2 = .16$. Pairwise comparisons confirmed the significant effect of relatedness in both modality conditions (auditory [$MD = 15.22$, $SE = 2.17$; $p < .001$]; visual [$MD = 7.15$, $SE = 2.11$; $p = .002$]), and showed that the interaction was due to a significantly greater effect of relatedness in the auditory condition compared to the visual condition. Most importantly, the results of the initial ANOVA demonstrated a clear effect of relatedness, and permitted the performance of a further analysis on responses to related trials.

4.2.2.1.2 Related trials only

A 2 x 2 within-subjects ANOVA examined the effects of modality and transformation (inverse, retrograde) when participants were responding to related targets only. The main effect of modality was once again significant, $F(1,52) = 5.59$, $MSE = .03$, $p = .022$, $\eta_p^2 = .10$, with more errors being made in the auditory ($M = 27.23$, $SD = 13.16$) compared to the visual condition ($M = 22.56$, $SD = 11.95$). There was also a significant main effect of transformation, $F(1,52) = 7.73$, $MSE = .03$, $p = .008$, $\eta_p^2 = .13$, with more errors being made when identifying inverse transformations ($M = 27.17$, $SD = 11.83$) compared to retrograde transformations ($M = 22.63$, $SD = 12.62$). There was a significant interaction between transformation and modality, $F(1,52) = 6.87$, $MSE = .03$, $p = .011$, $\eta_p^2 = .12$, which appeared to be due to a larger effect of transformation in the auditory condition (see Figure 4.4).

The 1½-D hypothesis predicted that in the auditory condition, lower PE would be observed in the inverse condition – but as can be seen from Figure 4.4 PE was actually lower in the retrograde condition. In contrast, the 2½-D hypothesis predicted that in the visual condition, there would either be no effect of transformation, or that PE would be lower in the retrograde condition. Pairwise comparisons were carried out to examine the simple effects of transformation, and confirmed that the effect in the auditory condition was highly significant *(MD* =

9.18, *SE* = 2.58; *p* = .001). In the visual condition, the mean difference between inverse and retrograde conditions was non-significant (*MD* = 0.09, *SE* = 2.23; *p* = .790). These results provided mixed support for the hypotheses – the 1½-D hypothesis was unsupported, but the 2½-D hypothesis was supported.

Further pairwise comparisons were carried out to examine the simple effects of modality in both transformation conditions. They revealed that, when identifying inverse transformations, there was a highly significant effect of modality (*MD* = 9.31, *SE* = 2.64; *p* = .001), with more errors being made in the auditory compared with the visual condition. On the other hand, there was no significant effect of modality when identifying retrograde transformations (*MD* = 0.03, *SE* = 2.43; *p* = .901). Participants were worse at identifying inverse transformations in the auditory compared with the visual condition. However, participants were able to identify retrograde transformations equally well in both modality conditions.

4.2.2.2 RT data

4.2.2.2.1 All trials

The purpose of analysing response time data was to obtain additional information about general trends in participants' performance, which might support or contradict the error data. Since no a priori theoretical criteria were applied to the relationship between PE and RT, the latter was analysed fully. Overall mean RT was 887.05ms (measured from the offset of the target). Mean results for responses to inverse, retrograde and unrelated targets in both modality conditions are displayed in Figure 4.5. Log-transformed RT data were subjected to the same analyses as were arcsine-transformed error data.

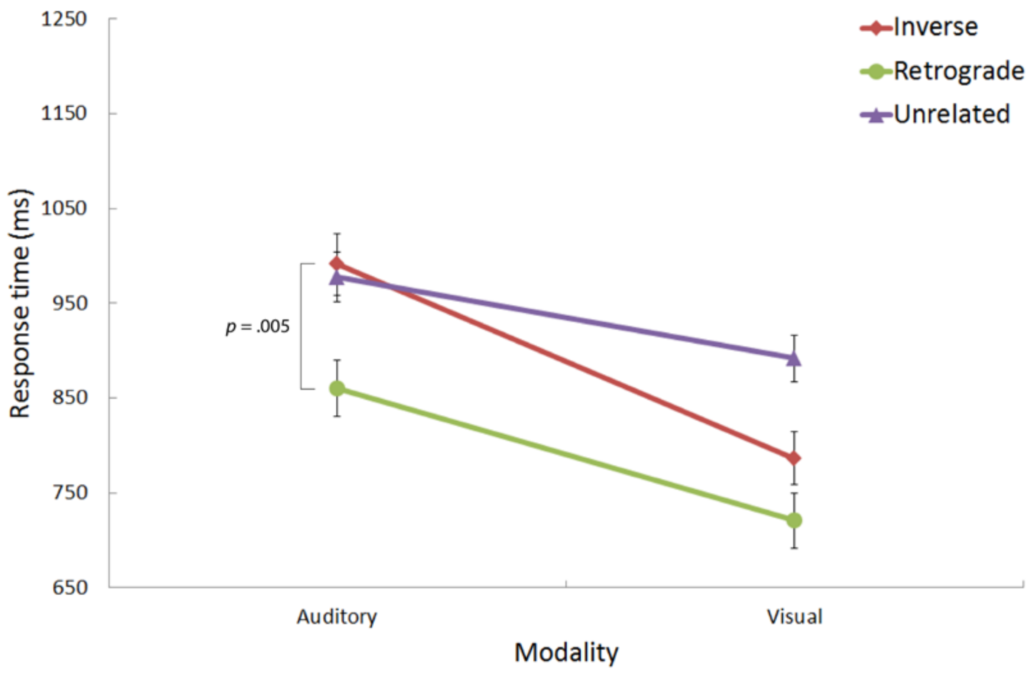


Figure 4.5. Experiment 2: Mean RT in target conditions, plotted as a function of modality. Significance values for simple effects of transformation were obtained from the analysis on log-transformed data (NB simple effects of modality are not displayed). Error bars indicate the standard error of the mean (SEM ± 1).

The 2 x 2 ANOVA, with the within-subjects factors relatedness and modality, revealed a highly significant main effect of relatedness, $F(1,52) = 18.15$, $MSE = .06$, $p < .001$, $\eta_p^2 = .26$, with slower responses to unrelated ($M = 934.75$, $SD = 344.18$) compared to related targets ($M = 839.35$, $SD = 305.34$). There was a highly significant main effect of modality, $F(1,52) = 13.07$, $MSE = .06$, $p = .001$, $\eta_p^2 = .20$, with slower responses in the auditory condition ($M = 951.53$, $SD = 369.35$) compared to the visual condition ($M = 822.58$, $SD = 290.27$). The interaction between relatedness and modality was also highly significant, $F(1,52) = 16.53$, $MSE = .02$, $p < .001$, $\eta_p^2 = .24$, and appeared to be due to there being a significantly greater effect of relatedness in the visual condition compared to the auditory condition. Paired comparisons revealed that the effect of relatedness was highly significant in the visual condition (*MD* = 138.22, *SE* = 31.22; $p < .001$) and approaching significance in the auditory condition (*MD* = 52.57, *SE* = 34.45; $p = .075$). In general, the pattern of results is in agreement with the error data findings (except with regard to the nature of the interaction between relatedness and modality): the conditions in which targets were identified most slowly, were the same conditions in which most errors were made. Importantly, this suggests that the RT data were not contaminated by a speed-accuracy trade-off (Pachella, 1974).

4.2.2.2.2 Related trials only

A further 2 x 2 ANOVA was carried out in order to examine the effects of transformation and modality for related targets only. The results confirmed the significant main effect of modality, $F(1,52) = 29.98$, $MSE = .07$, $p < .001$, $\eta_p^2 = .37$. Participants were slower to identify transformed targets in the auditory (*M* = 925.24, *SD* = 361.52) compared to the visual condition (*M* = 753.46, *SD* = 285.96). The main effect of transformation was also significant, $F(1,52) = 6.67$, $MSE = .07$, $p = .013$, $\eta_p^2 = .11$, with participants being slower to identify inverse (*M* = 888.62, *SD* = 364.45) compared to retrograde transformations (*M* = 790.09, *SD* = 285.99). Although visual inspection of Figure 4.5 suggests that, as with the error data, there was a greater difference between the transformation mean scores (inverse, retrograde) in the auditory condition compared to the visual condition, the interaction between transformation and modality merely approached significance, $F(1,52) = 2.83$, $MSE = .08$, $p = .098$, $\eta_p^2 = .05$.

Despite the failure of the interaction to reach (formal) significance, paired comparisons were carried out to more closely examine the simple effects in order to contrast with error data. The comparisons revealed a significant effect of transformation in the auditory condition (*MD* = 131.14, *SE* = 45.83; $p = .005$) but not in the visual condition (*MD* = 65.90, *SE* = 40.43; $p = .564$); in the auditory condition participants were slower to identify inverse compared to retrograde transformations, but in the visual condition participants identified both types of transformation equally quickly. Assuming that processing difficulty is reflected in response times, with slower responses indicating harder tasks, these results would appear to be in agreement with the results of the analysis on error data. Thus, the 1½-D hypothesis was once again unsupported, but the 2½-D hypothesis found some further support.

Finally, pairwise comparisons revealed a significant effect of modality in both transformation conditions (inverse [*MD* = 204.40, *SE* = 40.41; $p < .001$]; retrograde [*MD* = 139.16, *SE* = 44.06; $p = .013$]). In both instances, responses were slower when identifying transformations in the auditory condition.

4.2.2.3 Signal detection analysis

The results of the analysis on PE data may have been subject to response bias – depending on the condition, participants may have been more or less inclined to give a 'same' (i.e. 'yes, the target is related to the standard') versus a 'different' response (i.e. 'no, the target is unrelated to the standard'). As analysis of the PE data does not account for the effects of response bias when measuring detectability of the signal, further analysis was carried out using signal detection theory. The error data was converted to hits (correct responses to related targets) and false alarms (incorrect responses to unrelated targets) and d' and c measures were then calculated (Stanislaw & Todorov, 1999). d' is a measure of signal detectability that takes into account the contaminating effects of response bias to obtain cleaner results than PE alone. *c* is a measure of the bias.

4.2.2.3.1 Sensitivity to the signal

Overall *d'* was 1.06 (*SD* = 0.64) and the mean results in each condition are displayed in Figure 4.6. A 2 x 2 within-subjects ANOVA was carried out to examine the effects of modality (auditory, visual) and transformation (inverse, retrograde) on the detectability of related targets. The results mostly confirm those reported in the analysis of PE data. There were highly significant main effects of modality, $F(1,52) = 25.51$, $MSE = .42$, $p < .001$, $\eta_p^2 = .33$, with detectability being better in the visual ($M = 1.29$, $SE = .11$) compared to the auditory condition ($M = 0.83$, $SE = .09$), and of transformation, $F(1,52) = 17.43$, $MSE = .27$, $p < .001$, $\eta_p^2 = .25$, with retrograde transformations ($M = 1.21$, $SE = .10$) being more detectable than inverse transformations ($M = 0.91$, $SE = .09$).

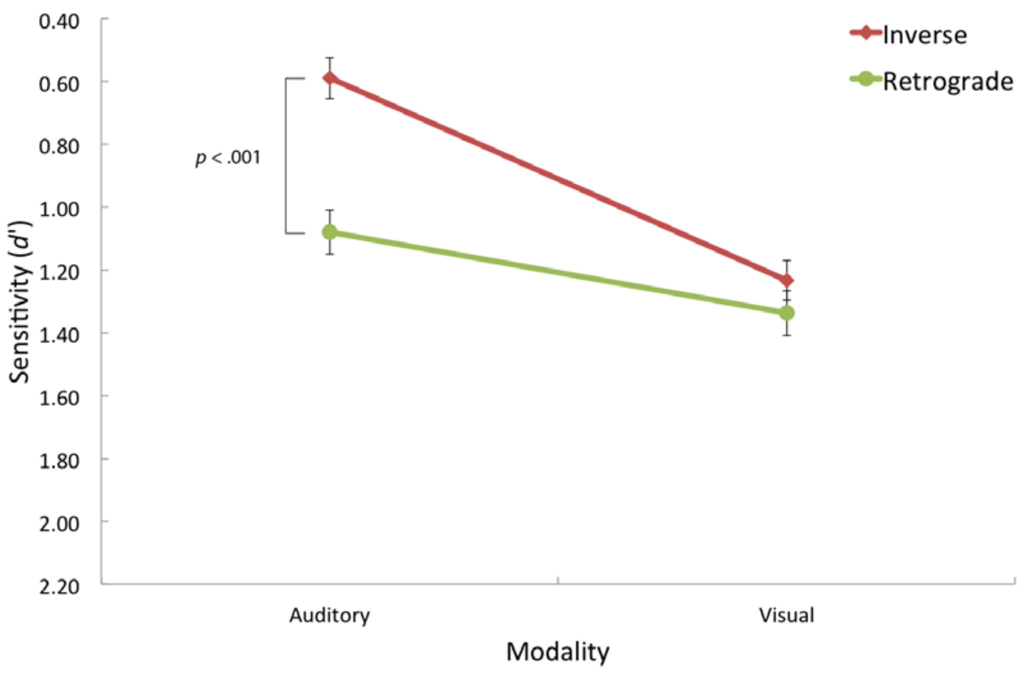


Figure 4.6. Experiment 2: Mean *d′* in transformation conditions, plotted as a function of modality (NB simple effects of modality are not displayed). The scale along the y-axis has been inverted so that results can be more easily compared with figures displaying PE data. Error bars indicate the standard error of the mean (SEM ± 1).

The interaction between transformation and modality was significant, $F(1,52) = 7.18$, $MSE = .28$, $p = .010$, $\eta_p^2 = .12$. Pairwise comparisons revealed that while the effect of transformation was highly significant in the auditory condition, with retrograde transformations being more detectable than inverse transformations ($MD = 0.49$, $SE = .10$; $p < .001$), there was no significant effect of transformation in the visual condition ($MD = 0.11$, $SE = .10$; $p = .306$). Thus, the results of the analysis on *d′* data also failed to support the 1½-D hypothesis but supported the 2½-D hypotheses.

Further comparisons revealed that the effect of modality was significant in both the retrograde ($MD = 0.26$, $SE = .12$; $p = .040$) and the inverse conditions ($MD = 0.64$, $SE = .11$; $p < .001$). In both conditions visual transformations were more detectable than auditory transformations. This was the only difference from the results of the analysis on PE, as in the PE analysis no effect of modality had been found on the recognition of retrograde transformations.

4.2.2.3.2 Response bias

Mean *c* was -0.15 (*SD* = 0.20). The negative value of *c* suggests that overall, participants were biased towards responding 'same' and were therefore more liberal than conservative with their responses. A 2 x 2 ANOVA was carried out to examine the effects of modality and transformation on response bias. There was a highly significant effect of modality, $F(1,52) = 13.99$, $MSE = .07$, $p < .001$, $\eta_p^2 = .21$, with responses being more liberal in the auditory condition ($M = -0.22$, $SE = .03$) compared with the visual condition ($M = -0.08$, $SE = .04$). In other words, participants were more biased towards indicating that targets were related to the standard in the auditory than in the visual condition. There were no further significant main effects or interactions.

4.2.2.4 Music training analysis

Finally, an exploratory analysis of the effects of music training was carried out. Many behavioural studies have demonstrated that participants with some music training perform better in melodic processing tasks than participants with no training (Halpern, Bartlett, & Dowling, 1998; Trainor, Desjardins, & Rockel, 1999). For example, Halpern, Bartlett and Dowling (1998) demonstrated that musicians are better at discriminating same/different melodies. It should be noted that the performance advantage associated with musical experience has been observed more strongly in tasks involving pitch interval perception rather than contour (Trainor et al., 1999). Surprisingly, previous behavioural studies that have investigated the processing of inverse and retrograde melodic transformations have not systematically investigated the effects of music training (Cupchik et al., 2001; Dowling, 1972; Krumhansl et al., 1987; McLachlan et al., 2010; Schulze et al., 2012): though they report levels of music training, they do not report the performance differences that may be associated with this variable. With regard to the possible effects in the visual condition of the present experiment, music training has also been previously linked with better performance in visuo-spatial tasks such as mental rotation (Pietsch & Jansen, 2012). For these reasons, it was anticipated that participants with some previous training might perform better in the auditory condition, and that this performance advantage may also transfer to the visual condition.

In order to examine the effects of music training on performance, participants were allocated to one of two groups based on their responses to the 'Demographic and Music Background Questionnaire'. 20 participants reported some previous music training and were allocated to the 'some training' group. 33 participants reported no previous music training and were allocated to the 'no training' group. Firstly, a Pearson's correlation was run to investigate any associations between the amount of music training in the 'some training' group, and performance in the different conditions of the experimental task (correlations were run on PE, RT, d' and c data). Only significant findings are reported here ($p \leq .05$) – the full results can be found in Appendix IV. There was a positive correlation between amount of music training and c data when discriminating inverse transformations in the visual condition, $r = .48$, $p = .032$ – as the amount of music training increased, so too did c scores. Mean c scores were general negative, so this means that increasing number of months/years music training was associated with decreasingly liberal responses in this condition. There were no further significant correlations.

Secondly, a series of 2 x 2 x 2 mixed ANOVAs, with the within-subjects factors modality (auditory, visual) and transformation (inverse, retrograde) and the between-subjects factor music training (some training, no training), were run on PE and RT data (related trials only), and on *d′* and *c* data. The results of interest were the main effect of music training and the interactions between music training and other factors. Any significant interactions were followed up with pairwise comparisons exploring the simple effects of music training. Once again, only significant findings are reported here. The full results of these analyses can be found in Appendix IV.

The analysis on PE data revealed a significant interaction between modality and music training, $F(1,51) = 5.59$, $MSE = .03$, $p = .022$, $\eta_p^2 = .10$. Pairwise comparisons were run to examine the simple effects of music training in both modality conditions. In the auditory condition, a significant difference was found between music training conditions ($MD = 9.61$, $SE = 3.52$; $p = .019$), with more errors being made by participants with no training ($M = 30.86$, $SD = 13.04$) compared to those with some training ($M = 21.25$, $SD = 11.28$). There were no further significant results. The analysis on RT data revealed a significant three-way interaction between modality, transformation and music training, $F(1,51) = 4.46$, $MSE = .07$, $p = .040$, $\eta_p^2 = .08$. However, pairwise comparisons that were run to examine the simple effects of music training in the different conditions failed to reveal any significant results. There were no further significant results revealed by the analysis on RT data. The analysis on d' and c data failed to reveal any significant results.

4.2.3 Discussion

The results of Experiment 2 revealed contrasting effects of transformation in the auditory and visual conditions, with performance in the auditory condition being better for targets under retrograde transformation, and no performance advantage for either transformation in the visual condition. These results failed to support the 1½-D hypothesis, but supported the 2½-D hypothesis.

The 2½-D hypothesis predicted that in the visual condition, inverse transformations would be recognised no better than retrograde transformations, and that retrograde transformations may be recognised more effectively than inverse transformations. This hypothesis was supported by all analyses of responses to visual targets, which demonstrated that there were no differences in performance for inverse and retrograde transformations. When interpreted within the SPS framework outlined in Chapter 2, this result may be explained by the relative compatibility of structural relations on two supramodal scalar dimensions. Visual stimuli that map the timing of events onto a horizontal dimension can be represented in a 2½-D supramodal pattern space. Vertical and horizontal spatial dimensions correspond to two scalar dimensions (accounting for two of the dimensions), and the timing of events corresponds to a temporal dimension (accounting for the ½ dimension). In order to recognise transformed targets, these representations would have to be mentally transformed. Both inverse and retrograde transformations could be achieved by inverting ordinal relations on an equivalent scalar dimension, explaining why there was no difference in performance for these transformations – they both involved the same process. However, this does not take into account the additional structural redundancy of representations on the temporal dimension, which prompted the prediction that retrograde transformations may be recognised more successfully than inverse transformations. Although the mean data demonstrates that performance was better for retrograde transformations, the effect of transformation failed to reach significance in all analyses of responses to visual patterns. As the results demonstrate there was no advantage for retrograde transformations, this would suggest that structural representations on the temporal dimension may have been discarded in favour of a more economical 2-D representation.

It should be noted that an alternative explanation for the results in the visual condition can be sought from the transformational approach to symmetry and similarity perception (Hahn, Chater, & Richardson, 2003; Hahn, 2014; Palmer, 1983). According to this approach, symmetry in two-dimensional representations of visual images is formalised as invariance under different types of transformation such as translation, rotation and reflection. In order to perceive symmetry, this approach implies that representations and mental transformations are analogous to real-world objects and manipulations in visual space. For this reason, the transformation process is typically attributed to mental rotation (Shepard & Metzler, 1971). When considering the visual stimulus in the present experiment as a static two-dimensional image, inverse and retrograde transformation would be equivalent to reflectional symmetry around different axes. In other words, the absence of a transformation effect in the visual condition could be interpreted as reflecting equivalent transformation processes that involve a mental rotation around the x-axis (inverse) or the y-axis (retrograde).

The explanation offered by the SPS framework has a clear affinity with the transformational approach, in that it also formalises structural regularity, or symmetry, as invariance under different types of transformation. However, the transformational approach is visuo-centric, and implies that the mechanism contributing to symmetry perception in visual objects is specific to the visual modality. The focus of the approach taken by the SPS framework is on a more abstracted, structural level, and explicitly states that the mechanism contributing to symmetry perception transcends specific sensory modalities. From the results reported in the present experiment it is impossible to draw any conclusions regarding the appropriateness of the interpretations offered by these different approaches.

The 1½-D hypothesis predicted that in the auditory condition inverse transformations would be processed more effectively than retrograde transformations. As with the prediction made by the 2½-D hypothesis, this prediction was based on the theoretical assumptions of the SPS framework, outlined in Chapter 2. It assumed that structural information, abstracted from auditory stimuli, would be represented in a 1½-D supramodal pattern space constructed from a scalar and a temporal dimension: the scalar dimension representing the relative pitch of tones and the temporal dimension representing the relative timing of tones. It was predicted that inverse transformations would be recognised more successfully because they require an inversion of ordinal relations on the scalar dimension, which is easier to process than the inversion of ordinal relations on the temporal dimension required to recognise retrograde transformations.

The finding that, to the contrary, retrograde transformations were recognised more successfully than inverse transformations would seem to call into question the assumptions of the SPS framework. For instance, perhaps the result indicates that inversions on temporal dimensions are actually easier to process than inversions on scalar dimensions. However, there is an issue with the way in which targets were presented that offers an alternative explanation for the unexpected transformation effect.

In the experiment, standard and target patterns were presented in the same pitch space (i.e. standard and target stimuli were composed from the same three pitches). As a result, when retrograde transformations were applied to target patterns, their tones preserved the pitches of standard tones, albeit in reverse order. In other words, retrograde-transformed targets preserved the non-structural, physical properties of the standard. Crucially, this redundant non-structural information could have been used to recognise that patterns were related, and thus facilitate performance. The same was not true for inverse transformations – the nature of the transformation meant that the same non-structural cues could not be used to identify related targets. Therefore, it could be argued that retrograde transformations had an unfair advantage because, whilst inverse transformations could only be recognised by processing structural information, retrograde transformations could be recognised by processing both structural and non-structural information. To investigate this possibility, further experiments would need to be carried out that eliminate the availability of non-structural information. This could be done by transposing targets to different pitch registers, or by presenting patterns in cross-modal trials.

The issue highlighted here – i.e. the availability of structural and non-structural information when perceiving pattern relationships – is not a trivial one. The recognition of relationships between patterns that share physical properties may engage sensory-specific mechanisms, in addition to any hypothesised supramodal mechanisms that are sensitive to structural information, and which are the focus of the present research. Yet this issue has not been properly addressed in the previous literature, which has inconsistently studied the perception of retrograde transformations either with (Dowling, 1972) or without additional transposition (Cupchik et al., 2001; Foster et al., 2013; Jones & Zamostny, 1975; Restle, 1976; Schulze et al., 2012; Zatorre et al., 2010). Indeed, a facilitation effect of non-structural information may explain the contradictory results reported by previous studies that have investigated the perception of inverse and retrograde transformations of melody. In a study by Dowling (1972) it was reported that inverse transformations were recognised more accurately than retrograde transformations, whilst Cupchik et al. (2001) reported that in their study retrograde transformations were recognised more accurately than inverse transformations. On inspection of their methodology, it is clear that they both applied retrograde transformations differently – Dowling transposed retrograde transformations, but Cupchik et al. did not. Thus, it can be concluded that the perception of pattern regularities described by retrograde transformation was facilitated when the transformation did not include a transposition. Although this data appears to support the explanation provided above (that redundant non-structural information facilitated the perception of pattern related under retrograde transformation), to date no research has been conducted that directly tests this interpretation.

Moving on from the issue of structural and non-structural information, an additional finding of interest was the performance advantage for visual patterns over auditory patterns. In general, participants were better at recognising transformed targets when they were presented visually. The visual advantage was less strong when recognising targets under retrograde transformation, but this may have been due to the fact that additional non-structural information cues could be used to recognise retrograde transformations in the auditory condition, but not in the visual condition. The visual advantage is consistent with previous research that has shown that analogous visual patterns are discriminated more accurately than auditory pitch patterns (Balch & Muscatelli, 1986). The present experiment extends these findings to show that this visual advantage persists at high levels of processing involving the recognition of pattern transformations. One interpretation of this might be that structural information is abstracted from visuo-spatial stimuli more efficiently than from auditory pitch stimuli. However, as has already been discussed, performance in the task may not have been based solely on the processing of structural information. Furthermore, according to the interpretation offered by the 2½-D hypothesis, structural representations abstracted from auditory and visual stimuli may not have been equivalent, and the visual advantage may reflect the fact that the mapping of timing onto the horizontal dimension provided additional redundancy.

Another interesting finding was revealed by the exploratory analysis on music training. Participants that had received some training performed better than those that had received no training, but only when recognising auditory patterns. This result is in agreement with previous research that has demonstrated that musicians are better at recognising melodies than non-musicians (Halpern et al., 1998; Trainor et al., 1999). However, it is not consistent with the idea that musical training would confer an advantage in visuo-spatial tasks (Pietsch & Jansen, 2012). It should be noted that any conclusions regarding these findings should be considered with caution – the analysis on music training was conducted post hoc, and as such the level of training was not carefully controlled. Participants' level of music training was determined by self-reports, and those placed in the 'some training' group had anywhere between less than 6 months and more than 8 years of music training. It is likely that any effects of music training on performance would be related to both to the number of years and type of training received. For example, participants with only a little training might have had a negligible performance advantage over participants with no previous training.

To conclude, the results of Experiment 2 supported the 2½-D hypothesis, but failed to support the 1½-D hypothesis. On the one hand, this may disprove the assumptions of the SPS framework. However, an important issue has been identified regarding the availability of non-structural information when retrograde targets are not transposed, which may have facilitated recognition (and obscured the effects of structural processing). One way to address this issue would be to investigate the recognition of transformations when patterns are presented in cross-modal conditions. This was done in Experiment 3.

4.3 Experiment 3: Cross-modal trials

The aim of Experiment 3 was to examine the hypotheses outlined at the beginning of this chapter in a cross-modal context (by presenting standard and target patterns in different sensory modalities). In each trial of the present experiment both an auditory and a visual pattern were presented. In half of the trials, an auditory standard was followed by a visual target (AV condition), and in the other half a visual standard was followed by an auditory target (VA condition).

As standard and target patterns were presented cross-modally, recognition could not be based on sensory specific, non-structural information. Instead, recognition could theoretically only be based on the processing of structural information. According to the SPS framework, auditory pitch patterns can be represented in a 1½-D supramodal pattern space. For patterns represented in such a space, the 1½-D hypothesis predicts that inverse transformations are easier to process than retrograde transformations. In contrast, it has been proposed that structural information abstracted from visual stimuli presented horizontally would be represented in a 2½-D supramodal pattern space. For patterns represented in such a space, the 2½-D hypothesis predicts that inverse transformations are no easier to process than retrograde transformations, and that retrograde transformations may even be processed more effectively (though the results from Experiment 2 supported the former).

The question is, in a cross-modal trial when an auditory standard must be compared with a visual target or a visual standard must be compared with an auditory target, how will participants identify transformations? Will they mentally transform structural representations abstracted from the standard, or will they mentally transform structural representations abstracted from the target? The answers to these questions are not immediately apparent, but have important implications for performance in the task, due to the contrasting supramodal representations hypothesised by the SPS framework: recognition performance for targets based on the mental transformation of patterns represented in a 1½-D pattern space will be different to recognition performance based on the mental transformation of patterns represented in a 2½-D pattern space.

The design of the experiment meant that participants knew on each trial which transformation will have been applied to related target patterns, so it was logical to assume that they would mentally transform standard patterns, in anticipation of the target. This assumption was supported by comments collected from participants after taking part in Experiment 2. If this were the case, then the competing 1½-D and 2½-D hypotheses could be allocated to modality conditions according to the modality of the standard – the 1½-D hypothesis would predict performance in the AV condition, and the 2½-D hypothesis would predict performance in the VA condition. However, it was impossible to know for certain which structural representation would be mentally transformed in each cross-modal condition. Therefore, both hypotheses were tested in both modality conditions. It was expected that contrasting effects of transformation would be observed in each modality condition, depending on the strategy used by participants (i.e. whether they mentally transformed representations of standard or target patterns).

In summary, the present experiment tested two competing hypotheses: the 1½-D and the 2½-D hypothesis. Support for the 1½-D hypothesis would imply that the recognition of transformed targets is based on the processing of structural information represented in a 1½-D supramodal pattern space. Support for the 2½-D hypothesis would imply that the recognition of transformed targets is based on the processing of structural information represented in a 2½-D supramodal pattern space. Although the hypotheses were not allocated to specific modality conditions, it was expected that if the results in one condition were best explained by the 1½-D hypothesis then the results in the other condition would be best explained by the 2½-D hypothesis.

4.3.1 Methods

4.3.1.1 Participants

31 students from the University of Roehampton took part in Experiment 3 (female = 16, male = 15; mean age = 25.55 years, *SD* = 8.26). They all had normal hearing and normal or corrected-to-normal vision. One participant reported they were left-handed and the remainder were right-handed. 20 participants (65%) reported some level of music training (mean = 4.78 years). They received course credit for their participation.

4.3.1.2 Stimuli

The stimuli used in Experiment 3 were identical in all aspects to those presented in Experiment 2.

4.3.1.3 Design and procedure

The design and procedure were identical to those of Experiment 2 with the following exceptions. Whereas in Experiment 2 each trial was unimodal (auditory or visual), in Experiment 3 each trial was cross-modal (auditory-visual [AV] or visual-auditory [VA]). Trials in the AV condition comprised an auditory standard that was followed by a visual target. Conversely, trials in the VA condition comprised a visual standard followed by an auditory target. This meant that the experimental training had to be adapted. Before each sub-block of the experiment, participants were informed that they would be comparing an auditory standard with a visual target or vice versa. Example stimuli and practice trials corresponded to the modality condition of each sub-block.

4.3.2 Results

Data from 2 participants were excluded from analysis because they failed to perform above chance levels on the task (their overall error rate was 50% or greater). As was performed in Experiment 2, paired-samples *t*-tests were run on block order to examine any effects of learning on PE (arcsine-transformed) and RT (log-transformed). Once again, both *t*-tests failed to reach significance (PE: $t(28) = 0.61$, $p = .275$, one-tailed; RT: $t(28) = -1.15$, $p = .130$, one-tailed].

4.3.2.1 Error data

4.3.2.1.1 All trials

Overall PE was 21.47, which was approximately 9% less than recorded in Experiment 2. It is possible that cross-modal presentation made the task easier, or that the accuracy advantage was due to a more skilled participant sample – the proportion of participants who had received some musical training was higher in the present experiment, and the mean number of years of music training that this group had received was also higher. However, an exploratory analysis of music training (reported below), failed to reveal any significant effects. Mean results for responses to retrograde, inverse and unrelated targets in both modality conditions are displayed in Figure 4.7.

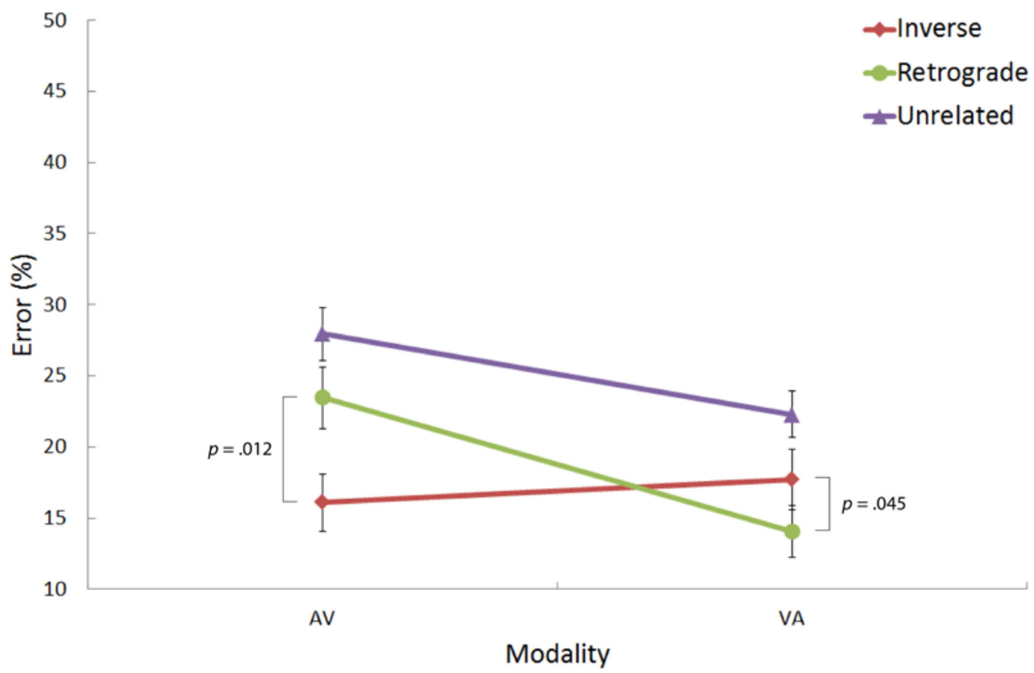


Figure 4.7. Experiment 3: Mean PE in target conditions, plotted as a function of modality. Significance values for simple effects were obtained from the analysis on arcsine-transformed data (NB simple effects of modality are not displayed). Error bars indicate the standard error of the mean (SEM ± 1).

A 2 x 2 ANOVA with relatedness and modality as the within-subjects factors was run on arcsine-transformed PE data. There was a significant main effect of relatedness, $F(1,28) = 5.46$, $MSE = .03$, $p = .027$, $\eta_p^2 = .16$, with more errors being made when identifying unrelated ($M = 25.12$, $SD = 16.41$) compared with related targets ($M = 17.82$, $SD = 11.58$). The main effect of modality was also significant, $F(1,28) = 10.68$, $MSE = .02$, $p = .003$, $\eta_p^2 = .28$, with more errors in the AV condition ($M = 23.85$, $SD = 12.04$) compared with the VA condition (*M*

= 19.08, *SD* = 14.72). The interaction between relatedness and modality failed to reach significance.

4.3.2.1.2 Related trials only

A further 2 x 2 repeated-measures ANOVA was carried out in order to examine the effects of transformation and modality when responding to related targets only. There was a significant main effect of modality, $F(1,28) = 4.88$, $MSE = .03$, $p = .035$, $\eta_p^2 = .15$, with more errors being made in the AV ($M = 19.77$, $SD = 13.30$) compared with the VA condition ($M = 15.86$, $SD = 12.27$). The main effect of transformation failed to reach significance. However, the interaction between transformation and modality was highly significant, $F(1,28) = 13.40$, $MSE = .02$, $p = .001$, $\eta_p^2 = .32$. Visual inspection of Figure 4.7 suggests that the interaction was due to the effect of transformation being reversed across modality conditions. In the AV condition participants made more errors when identifying retrograde transformations, but in the VA condition participants made more errors when identifying inverse transformations. Paired comparisons revealed that the effect of transformation was significant in both the AV condition ($MD = 7.36$, $SE = 2.55$; $p = .012$) and the VA condition ($MD = 3.68$, $SE = 2.72$; $p = .045$). The results in the AV condition are best explained by the 1½-D hypothesis, whilst the results in the VA condition are best explained by the 2½-D hypothesis. This would seem to suggest that recognition was based on the processing of structural representations abstracted from the standard stimulus.

Further pairwise comparisons were carried out to examine the simple effects of modality in both transformation conditions. They revealed that, when identifying retrograde targets, participants made more errors in the AV condition compared with the VA condition (*MD* = 9.43, *SE* = 2.42; $p < .001$). However, there was no effect of modality when identifying inverse transformations (*MD* = 1.61, *SE* = 2.70; $p = .575$). Participants were able to identify inverse transformations equally well in both modality conditions.

4.3.2.2 RT data

4.3.2.2.1 All trials

Although mean PE in Experiment 3 was substantially lower compared to Experiment 2, mean RT was approximately 180ms slower (*M* = 1065.60ms) – therefore, targets in the cross-modal trials were recognised more accurately but responses were slower. Mean results for responses to inverse, retrograde and unrelated targets in both modality conditions are displayed in Figure 4.8.

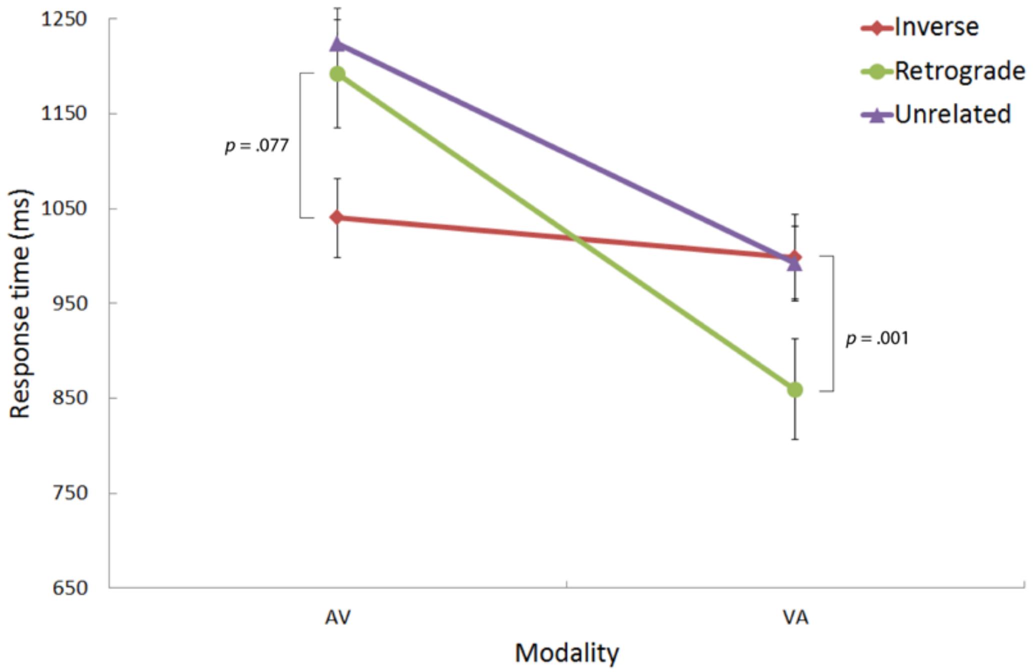


Figure 4.8. Experiment 3: Mean RT in target conditions, plotted as a function of modality. Significance values for simple effects were obtained from the analysis on log-transformed data (NB simple effects of modality are not displayed). Error bars indicate the standard error of the mean (SEM ± 1).

An initial 2 x 2 ANOVA, with relatedness and modality as the within-subjects factors, revealed a significant main effect of relatedness, $F(1,28) = 6.76$, $MSE = .06$, $p = .015$, $\eta_p^2 = .20$, with slower responses to unrelated targets (M = 1108.68, SD = 400.27) compared to related targets (M = 1022.52, SD = 399.04). The interaction between relatedness and modality failed to reach significance. There was a highly significant main effect of modality, $F(1,28) = 20.67$, MSE = .07, $p < .001$, $\eta_p^2 = .43$, with slower responses in the AV (M = 1170.37, SD = 426.16) compared to the VA condition (M = 960.83, SD = 398.52). These results are mostly in agreement with the error data.

4.3.2.2.2 Related trials only

A further 2 x 2 ANOVA examined the effects of transformation and modality on RTs for related targets only. The main effect of modality was significant, $F(1,28) = 9.83$, $MSE = .11$, $p = .004$, $\eta_p^2 = .26$, with slower RTs in the AV ($M = 1116.20$, $SD = 438.27$) compared to the VA condition ($M = 928.85$, $SD = 439.56$). The main effect of transformation failed to reach significance, $F(1,28) = 0.70$, $MSE = .09$, $p = .411$, $\eta_p^2 = .02$. However, there was a highly significant interaction between transformation and modality, $F(1,28) = 15.80$, $MSE = .07$, $p < .001$, $\eta_p^2 = .36$. Visual inspection of Figure 4.8 suggests that the interaction was due to a reversal of the transformation effect, with participants slower to identify retrograde compared with inverse transformations in the AV condition, but slower to identify inverse compared with retrograde transformations in the VA condition. Pairwise comparisons revealed that the effect of transformation in the AV condition was approaching significance ($MD = 151.89$, $SE = 68.08$; $p = .077$), whilst in the VA condition it was highly significant ($MD = 139.03$, $SE = 47.71$; $p = .001$). In the AV condition, the direction of the effect is best explained by the 1½-D hypothesis, which predicted that inverse transformations would be recognised more effectively than retrograde transformations. However, as the effect failed to reach formal significance, support for this hypothesis should be accepted with some caution. Taking a conservative approach, this result can also be explained by the 2½-D hypothesis, which predicted that recognition for inverse transformations would be no better for inverse transformations than for retrograde transformations. The 2½-D hypothesis also predicted that retrograde transformations could be recognised more effectively than inverse transformations. Therefore, results in the VA condition can be best explained by the 2½-D hypothesis. These results are largely in line with those revealed by the analysis on PE data. They also suggest that recognition was based on the processing of structural information abstracted from the standard.

Further pairwise comparisons revealed that the effect of modality on responses to retrograde transformations was highly significant (*MD* = 332.81, *SE* = 92.92; $p < .001$), with slower responses in the AV compared with the VA condition. On the other hand, there was no significant effect of modality for responses to inverse transformations (*MD* = 41.89, *SE* = 66.88; $p = .895$). When identifying inverse transformations, there was no difference in participants' response across modality conditions. This pattern of results mirrors that found in the error analysis.

4.3.2.3 Signal detection analysis

Further analysis was carried out using signal detection theory.

4.3.2.3.1 Sensitivity to the signal

The analysis on *d'* was largely in agreement with the analysis on PE data. Overall *d'* was 1.69 (*SD* = 0.90), which was greater than the overall *d'* in Experiment 2 – targets were more detectable in the cross-modal trials compared with the unimodal trials. Mean results in each condition are displayed in Figure 4.9.

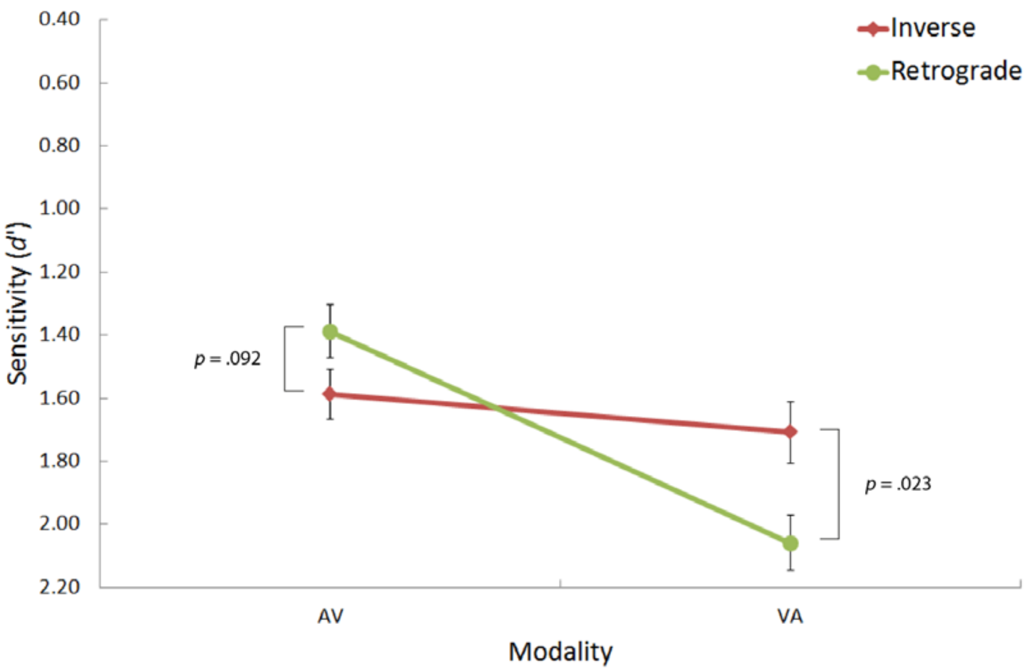


Figure 4.9. Experiment 3: Mean *d'* in transformation conditions, plotted as a function of modality (NB simple effects of modality are not displayed). The scale along the y-axis has been inverted so that results can be more easily compared with figures displaying PE data. Error bars indicate the standard error of the mean (SEM ± 1).

The 2 x 2 ANOVA revealed a significant main effect of modality, $F(1,28) = 11.46$, $MSE = .40$, $p = .002$, $\eta_p^2 = .29$, with detectability being better in the VA condition ($M = 1.88$, $SE = .20$) compared to the AV condition ($M = 1.49$, $SE = .16$). The main effect of transformation failed to reach significance. There was a highly significant interaction between transformation and modality, $F(1,28) = 12.57$, $MSE = .17$, $p = .001$, $\eta_p^2 = .31$. Pairwise comparisons revealed that the effect of transformation approached significance in the AV condition ($MD = 0.20$, $SE = .11$; $p = .092$), with inverse transformations being detected more easily than retrograde transformations. The effect of transformation was significant in the VA condition ($MD = 0.35$, $SE = .15$; $p = .023$), with retrograde transformations being detected more easily than inverse transformations.

The direction of the transformation effect in the AV condition is once again best explained by the 1½-D hypothesis. However, this explanation is accepted with some caution, as the effect failed to reach formal significance. Taking a conservative approach, the absence of a transformation effect may be better explained by the 2½-D hypothesis, which predicted that performance should be no better for inverse than for retrograde transformations. The results in the VA condition are best explained by the 2½-D hypothesis, as it also predicted that retrograde transformations might be processed more effectively than inverse transformations.

Further comparisons revealed that the effect of modality was highly significant in the retrograde condition (*MD* = 0.67, *SE* = .14; $p < .001$), with targets being better detected in the VA compared with the AV condition. There was no significant effect of modality when detecting inverse transformations (*MD* = 0.12, *SE* = .14; $p = .396$).

4.3.2.3.2 Response bias

Mean *c* was -0.09 (*SD* = 0.24). This was less negative than the overall *c* value in Experiment 2, suggesting that participants were less biased towards 'same' responses, though they assumed a liberal criterion. The 2 x 2 ANOVA revealed no significant main effects of transformation or modality, but the interaction between transformation and modality was approaching significance, $F(1,28) = 3.04$, $MSE = .06$, $p = .092$, $\eta_p^2 = .10$. As no hypotheses had been made regarding response bias, the interaction was not examined any further.

4.3.2.4 Music training analysis

An exploratory analysis was carried out to examine any effects of music training on performance (see Appendix IV for the results from all analyses). 18 participants were allocated to the some training group, and 11 participants were allocated to the no training group. A Pearson's correlation analysis revealed a significant positive correlation between amount of music training and *c* data when discriminating retrograde transformations in the VA condition, $r = .61$, $p = .007$. This indicated that increasing amount of music training was associated with decreasingly liberal responses in this condition. There were no further significant correlations. A series of 2 x 2 x 2 mixed ANOVAs, with the within-subjects factors modality (AV, VA) and transformation (inverse, retrograde) and the between-subjects factor music training (some training, no training), were run on PE, RT, *d'* and *c* data. All analyses failed to reveal any significant results.

4.3.3 Discussion

The results of Experiment 3 revealed contrasting effects of transformation in different modality conditions. In the AV condition the pattern of results were best explained by the 1½-D hypothesis, which predicted that inverse transformations would be processed more effectively than retrograde transformations. This would suggest that, in the AV condition, target patterns were identified by mentally transforming representations of auditory standard patterns. However, this interpretation is made with some caution as the effect of transformation only reached formal significance when PE data were analysed. In the VA condition, the opposite effect of transformation was found in all analyses – performance was better for targets under retrograde transformation. This result is best explained by the 2½-D hypothesis, which predicted that there would be no performance advantage for inverse transformations, and that retrograde transformations may be processed more effectively than inverse transformations. This would suggest that, in the VA condition, target patterns were recognised by mentally transforming representations of visual standard patterns.

Thus, the results of Experiment 3 provide some support for the SPS framework, based on the assumption that performance patterns reflect mental processes that operated on representations constructed from the standard patterns. Participants' comments, collected after they took part in the experiment, are in agreement with this assumption – they reported trying to anticipate the related target by applying the relevant mental transformation to the standard pattern. In the AV condition, the standard was an auditory pitch pattern. It has been proposed that structural information, abstracted from auditory pitch patterns, can be represented in a 1½-D supramodal pattern space, constructed from a scalar (representing relative pitch) and a temporal dimension (representing the relative timing of tones). In order to recognise an inverse transformation, ordinal relations must be inverted on the scalar dimension, and to recognise a retrograde transformation ordinal relations must be inverted on the temporal dimension. It has been hypothesised that inversions on a scalar dimension are processed more easily than inversions on a temporal dimension, due to the latter dimension's inherent directionality. This hypothesis was supported by the results in the AV condition, which demonstrated a partial performance advantage for inverse transformations.

Analysis of the PE data revealed that performance was significantly better for inverse transformations. Although the analysis on RT data failed to reach formal significance, the pattern of results was in the direction predicted by the 1½-D hypothesis – mean RT was faster when identifying inverse transformations. The same was true for the analysis on *d′* data. Although the effect of transformation failed to reach formal significance, the pattern of results was also in the direction predicted by the 1½-D hypothesis – mean detection rates were better for inverse than retrograde transformations. Nevertheless, this raises the question, why was the significant effect observed in the analysis of PE data not also observed in the analysis on *d′* data? Measures of signal sensitivity such as *d′* take into account both correct responses to related targets, i.e. hits, and incorrect responses to unrelated targets, i.e. false alarms. On inspection of the data, it appears that although the proportion of hits was higher for inverse transformations (inverse = 0.83, retrograde = 0.75), the proportion of false alarms was no better than for retrograde transformations (inverse = 0.28, retrograde = 0.28). In other words, participants were just as likely to mistakenly accept unrelated targets as inverse transformations as they were to accept unrelated targets as retrograde transformations. This may account for the absence of an effect of transformation on signal sensitivity data.

In the VA condition the results were clear – performance was better for targets under retrograde transformation. This result could not be explained by the 1½-D hypothesis, and was better explained by the 2½-D hypothesis. According to the 2½-D hypothesis, the visual stimuli used in the present experiment can be represented in a 2½-D supramodal pattern space, constructed from two scalar dimensions (corresponding to the vertical and horizontal visual dimensions) and a temporal dimension (corresponding to the timing of visual events). Whereas inverse and retrograde transformations of patterns represent in a 1½-D space involve inversions of ordinal relations on different types of supramodal dimensions (scalar or temporal), the same transformations of pattern represented in a 2½-D space can be achieved by inverting ordinal relations on the same type of supramodal dimension (scalar). This would imply that inverse transformations are equivalent to retrograde transformations, and that no performance advantage for inverse transformations should be observed. Furthermore, the additional structural redundancy represented on the temporal dimension predicted that performance may even be better for retrograde transformations. This hypothesis appeared to be supported by the results in the VA condition, which demonstrated a clear performance advantage for retrograde transformations.

One issue with this interpretation concerns the fact that in the VA condition of the present experiment, an advantage for retrograde transformations was observed, whilst in Experiment 2 no advantage was observed for either transformation, and yet both results have been interpreted as providing support for the 2½-D hypothesis. Why should the additional structural redundancy on the temporal dimension have facilitated retrograde transformations in the cross-modal task, but not in the unimodal task? The answer to this question is not immediately clear. In Section 4.2.3 it was suggested that the absence of a transformation effect in Experiment 2 was due to structural information on the temporal dimension being discarded, meaning that mental transformations were performed on 2-D representations constructed from two scalar dimensions. It is possible that in the present cross-modal experiment, structural information on the temporal dimension was not discarded because transformed patterns had to be compared with auditory pitch patterns, which could not be represented on two scalar dimensions. Another possible explanation may be sought in strategies involving visual and auditory mental imagery. Participants commenting on their experience of the experimental task frequently reported 'seeing' the auditory patterns presented in cross-modal trials, and less frequently reported 'hearing' the visual patterns. The mental imagery of visual objects has been well researched (e.g. Kosslyn, 1980), and it is reasonable to assume that participants were able to visualise the auditory stimuli. In the VA condition, target patterns under retrograde transformation may have been recognised by simply holding a 2-D representation of the visual standard in memory, visualising the relative vertical heights of the auditory target's tones, and retracing the visual standard from right to left as the tones of the auditory target were presented. The same strategy could not have been used to recognise inverse transformations in the VA condition, which must have required the mental transformation of structural information.

Mental images of sounds can also be produced from visual stimuli. As with visual images, auditory images possess a sensory quality that makes the experience of imagining sound similar to that of perceiving it (Zatorre, Halpern, Perry, Meyer, & Evans, 1996; Zatorre & Halpern, 1993). Gordon (1975) has called the internal analogue of aural perception 'audiation'. If participants were able to audiate visual targets in the AV condition as easily as they were able to visualise auditory targets in the VA condition, the pattern of results should have been similar to those that were found in the unimodal auditory condition of Experiment 2. However, this was not the case. Whereas in the auditory condition of Experiment 2 performance was better for retrograde transformations, in the AV condition of Experiment 3 performance was better for inverse transformations. This would suggest that participants were not able to audiate visual targets as effectively as they might have been able to visualise auditory targets in the VA condition. If this was the case, then recognition in the AV condition would have required the mental transformation of structural information. Following the same line of thought, however, visualisation may also have influenced responses in the AV condition. If some participants had visualised auditory standards, then structural representations may have been constructed on an additional scalar dimension (corresponding to the visualised horizontal dimension), which might explain why the effect of transformation was only marginal.

Finally, it should be noted that all analyses revealed consistent effects of modality. In general, recognition was better in the VA condition when visual standards were followed by auditory targets. When viewed in context of the visual advantage observed in Experiment 2, this is consistent with the assumption that recognition was based on the processing of standard stimuli. In other words, the visual advantage observed in Experiment 2 was expressed in the VA condition of Experiment 3 because target recognition was based on the processing of structural information, abstracted from the standard stimulus. Interestingly, there was no effect of modality on responses to inverse transformations. Interpreted within the SPS framework, this suggests that the general modality effect may not have resulted from structural information being abstracted more efficiently from visual stimuli (as suggested in Section 4.2.3 to explain the observed visual advantage), but instead have resulted from the fact that auditory and visual stimuli were represented in 1½- or 2½-D supramodal pattern spaces, respectively. Whilst inverse transformations in both modality conditions required an inversion of ordinal relations on a scalar dimension (and hence were equally effective), retrograde transformations were easier in the VA condition because they required an inversion on a scalar dimension, whereas in the AV condition they required an inversion on a temporal dimension, which is harder to process (due to the dimension's directionality). With this in mind, it would be interesting to see whether the modality effect disappears when auditory and visual stimuli are treated more equally, and can both be represented in a 1½-D supramodal pattern space.

In conclusion, the results from Experiment 2 provided some support for the SPS framework, outlined in Chapter 2. Assuming recognition responses were based on the mental transformation of structural information abstracted from standard stimuli, results in the AV condition were partially explained by the 1½-D hypothesis, and results in the VA condition were best explained by the 2½-D hypothesis. However, it is possible that the horizontal presentation of visual stimuli allowed participants to adopt a strategy that did not require the mental transformation of structural information. Further experiments were required in which auditory and visual stimuli are treated more equally, and can both be represented in a hypothesised 1½-D supramodal pattern space. Hence, Experiment 5 replicated the present experiment using visual stimuli that do not map time onto the horizontal dimension, and present all temporal events at a single location.

4.4 General discussion

The findings from Experiments 2 and 3 (summarised in Table 4.1) provided some initial support for the SPS framework outlined in Chapter 2. According to the SPS framework, structural information, abstracted from auditory and visual stimuli, is represented in a supramodal pattern space constructed from one or a combination of two qualitatively distinct dimensions: a scalar and a temporal dimension. The way in which this pattern space is constructed is dependent on the stimulus. One of the assumptions of the SPS framework is that mental transformations that require inversions of ordinal relations on a temporal dimension are harder to process than those that require inversions of ordinal relations on a scalar dimension, due to the former dimension's inherent directionality.

Table 4.1

Chapter 4: Summary of experiments, hypotheses and results

Experiment	Modality	Hypothesis tested	Result
2	A	1½-D	Not supported
	V	2½-D	Supported
3	AV	1½-D and 2½-D	Some support for 1½-D hypothesis
	VA	1½-D and 2½-D	Support for 2½-D hypothesis

Note. A = auditory; V = visual; AV = auditory-visual; VA = visual-auditory

An auditory pitch pattern can be represented in a 1½-D supramodal pattern space, constructed from a scalar and a temporal dimension – the scalar dimension represents relative pitch height and the temporal dimension represents the relative timing of tones. As retrograde transformations of patterns represented in a 1½-D space require inversions on a temporal dimension and inverse transformations require inversions on a scalar dimension, the 1½-D hypothesis predicted that targets under inverse transformation would be recognised more successfully than those under retrograde transformation.

The 1½-D hypothesis was unsupported in Experiment 2, as in the auditory condition a processing advantage for retrograde transformations was found. This might mean that inversions of ordinal relations are actually processed more effectively on temporal dimensions. However, it was noted that, due to the way in which stimuli were presented, the recognition of retrograde transformations could be based on the processing of structural information and additional non-structural information, but inverse transformations could only be based on the processing of structural information. Thus, it was possible that redundant non-structural information in the retrograde transformation condition facilitated recognition, obscuring any results that reflected structural processing. This interpretation found some support in Experiment 3 – when patterns were presented in cross-modal trials, and hence recognition could not have been based on non-structural cues, some support was found for the 1½-D hypothesis in the AV condition. It was reasoned that this was because target recognition was based on the processing of structural information abstracted from the auditory standard, which was represented in a 1½-D supramodal pattern space.

Informed by previous research that has also made a structural analogy between auditory pitch and visuo-spatial patterns, the visual stimuli used in experiments were presented sequentially and mapped auditory pitch onto vertical height. In addition, the timing of events was mapped onto the horizontal dimension. According to the SPS framework, stimuli presented in this way can be represented in a 2½-D supramodal pattern space, constructed from two scalar dimensions and a temporal dimension. The two scalar dimensions correspond to the relative vertical and horizontal positions of objects, whilst the temporal dimension corresponds to the relative timing of visual events. Inverse and retrograde transformations of patterns represented in a 2½-D space could both be achieved by inverting ordinal relations on a scalar dimension, and hence were (in theory) equally effective. Thus, an alternative 2½-D hypothesis was generated which predicted that, in contrast to the 1½-D hypothesis, inverse transformations should be recognised no more successfully than retrograde transformations. Furthermore, it predicted that retrograde transformations may even be recognised more successfully, due to the additional structural redundancy represented on the temporal dimension.

Support for the 2½-D hypothesis was found in Experiment 2 and Experiment 3. In Experiment 2, no effect of transformation was observed for inverse and retrograde transformations when visual standard and target patterns had to be compared. It was suggested that retrograde transformations were not recognised any more successfully than inverse transformations because structural representations on the temporal dimension were discarded. In Experiment 3, when visual standards were followed by auditory targets, retrograde transformations were recognised more successfully than inverse transformations. The performance advantage for retrograde transformations in this context was attributed to the retention of structural information on the temporal dimension, or to alternative strategies that might have been used by participants that did not require mental transformation of structural information.

Despite the support found for the SPS framework, it was difficult to draw any definitive conclusions from these first two experiments about the possibility of shared cognitive mechanisms being involved in the mental transformation of sequential pattern structure. Due to the way in which stimuli were presented, auditory and visual stimuli corresponded to different representations in the hypothetical supramodal pattern space. Furthermore, although in most conditions recognition was assumed to be based on the processing of structural information and in theory engaged the supramodal mechanisms under investigation, in one condition recognition could also be based on non-structural information and might therefore also have engaged other sensory specific mechanisms. In order to address these issues, further experiments are required that treat auditory and visual stimuli more equally (i.e. they can both be represented in a 1½-D supramodal pattern space), and that ensure recognition can only be based on the processing of structural information. When compared with the results of the experiments reported in the present chapter, this would provide a better examination of the SPS framework. The experiments reported in Chapter 5 attempted to do this.

Chapter 5: Transformation recognition in equivalent supramodal pattern spaces

5.1 Introduction

The general aim of Experiments 4, 5 and 6 was the same as Experiments 2 and 3 – to explore the possibility that shared cognitive mechanisms are involved in the mental transformation of sequential pattern structure, within the SPS framework outlined in Chapter 2. As was the case with the previous experiments, the investigation focussed on the processing of auditory pitch patterns and analogous visuo-spatial patterns that had undergone one of two types of isomorphic transformation: inverse and retrograde. Unlike in previous experiments, auditory and visual patterns were dimensionally equivalent, and could both be represented in a 1½-D supramodal pattern space.

The SPS framework proposes that auditory and visual stimuli can be represented in a supramodal pattern space, the construction of which depends on the stimulus. Auditory pitch patterns are represented in a 1½-D supramodal pattern space, constructed from a scalar dimension (representing the relative pitch of tones) and a temporal dimension (representing the relative timing of tones). One of the assumptions of the SPS framework is that inversions of ordinal relations on the temporal dimension are harder to process than inversions of ordinal relations on the scalar dimension, due to the former dimension's inherent directionality. For patterns represented in a 1½-D space inverse transformations require inversions on a scalar dimension, whilst retrograde transformations require inversions on a temporal dimension. In turn, this predicts that retrograde transformations of auditory pitch patterns should be harder to process than inverse transformations. This 1½-D hypothesis found some support in Experiment 3.

The experiments reported in Chapter 4 were unable to test this hypothesis in conditions when recognition was based on the processing of visual stimuli because they were presented in such a way that their structural abstractions were represented in a different pattern space (2½-D). The experiments reported in the present chapter sought to address this issue by presenting visual stimuli that could be represented in the same supramodal pattern space as auditory stimuli. In order to achieve this, the horizontal mapping of time was discarded from visual stimuli, so that all events were presented sequentially at different vertical heights but at the same central location on the horizontal axis. Theoretically, visual stimuli presented in this way would also be represented in a 1½-D supramodal pattern space, constructed from a scalar dimension (representing the relative vertical position of visual objects) and a temporal dimension (representing the relative timing of visual objects).

In order to test the 1½-D hypothesis the experiments reported in the present chapter used the same short-term recognition paradigm that was used in Chapter 4. Experiments 4 and 5 replicated Experiments 2 and 3, replacing horizontally presented with vertically presented visual stimuli. Experiment 6 was a hybrid experiment that combined unimodal and cross-modal conditions by presenting standard patterns in a single modality (auditory or visual), but target patterns bimodally (auditory and visual stimuli were presented simultaneously). Experiment 6 also applied transformations to targets with additional transposition.

5.2 Experiment 4: Unimodal trials

Experiment 4 replicated Experiment 2, replacing horizontally presented visual stimuli with centrally presented visual stimuli. By presenting visual stimuli centrally, they corresponded to a 1½-D supramodal pattern space, and thus the auditory and visual patterns used in the experiment were, in theory, dimensionally equivalent. The 1½-D hypothesis, which predicted that inverse transformations would be recognised more successfully than retrograde transformations, was tested in both modality conditions.

In addition to altering the way in which visual patterns were presented, the modality condition was changed from within-subjects to between-subjects. Within-subjects designs can be useful because they eliminate individual differences between the experimental conditions. However, one problem with this design in the present context is that it makes explicit the analogy between auditory and visual stimuli. This is problematic for two reasons: firstly, it encourages participants to treat auditory and visual patterns similarly, which may bias the results; secondly, it might encourage participants to 'visualise' auditory stimuli or 'audiate' visual stimuli, which was highlighted as a potential issue in the discussion of the previously reported experiments. Therefore, different groups of participants took part in the auditory and visual conditions, making modality a between-subjects factor. It was hoped that in making modality a between-subjects factor the experiment would, at least in some ways, be more rigorous.

5.2.1 Methods

5.2.1.1 *Participants*

58 students from the University of Roehampton took part in Experiment 4 (female = 45, male = 13; mean age = 23.16 years, *SD* = 9.90). One participant reported minor hearing problems, but this did not interfere with their ability to perform the task. All participants had normal or corrected-to-normal vision. Three participants were left-handed, one was ambidextrous, and the remainder were right-handed. 28 participants (48%) reported some previous music training (mean = 4.38 years). They all received course credit for their participation.

5.2.1.2 *Stimuli*

Figure 5.1 shows a sample pattern structure and its corresponding auditory and visual realisations. The stimuli used in Experiment 4 were identical in all aspects to those presented in Experiment 2 and 3, except with respect to the duration of each sequence event and the horizontal presentation of visual stimuli. Whereas visual stimuli in Experiments 2 and 3 were presented in sequential order from left to right on the display screen, the decision was made to remove the horizontal dimension in Experiment 4.

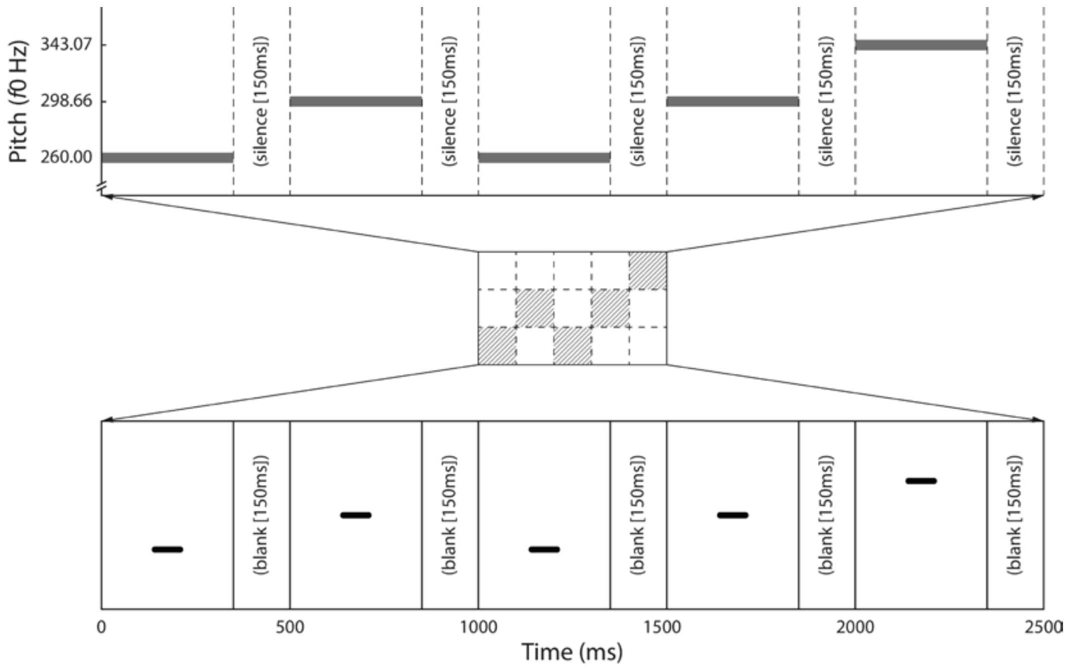


Figure 5.1. Auditory and visual stimuli used in Experiment 4. Top: A time-frequency plot of an auditory stimulus. An auditory stimulus consisted of a sequence of 5 tones (350ms durations) presented at one of 3 different pitch heights. Bottom: Each panel displays an image 'still' that was presented for 350ms. A visual stimulus consisted of a sequence of black bar segments presented at one of 3 different vertical heights.

As a result, each visual stimulus consisted of a sequence of segments presented at different vertical heights at the horizontal centre of the screen. An ISI between the offset and onset of segments had to be introduced because otherwise successive segments presented at the same vertical height appeared as a single object. The duration of each event was thus reduced from 500ms to 350ms whilst the IOI between events was held constant at 500ms. This created a 150ms ISI between the offset and onset of segments and ensured that successive segments that appeared at the same height were perceived as distinct events. In the interest of preserving the close analogy between visual patterns and auditory patterns, the duration of auditory tones was also reduced to 350ms to create a 150ms ISI between their offset and onset.

5.2.1.3 *Design and procedure*

The design and procedure were identical to those of Experiments 2 and 3 with the exception that modality was made a between-subjects factor rather than a within-subjects factor. As with Experiments 2 and 3, there were 6 experimental conditions – two levels of modality (auditory, visual), two levels of relatedness (related, unrelated), and two levels of transformation (retrograde, inverse) embedded into the former relatedness level. Consequently, the proportion of trials per condition was as follows: 1) auditory, related, retrograde (ARR) = 12.5%; 2) auditory, related, inverse (ARI) = 12.5%; 3) auditory, unrelated (AU) = 25%; 4) visual, related, retrograde (VRR) = 12.5%; 5) visual, related, inverse (VRI) = 12.5%; 6) visual, unrelated (VU) = 25%.

The same 15 standard patterns that were selected for use in Experiments 2 and 3 were used. Each pattern was presented once in related conditions and twice in unrelated conditions, making a total of 120 trials. Trials were divided by modality into two experimental sessions of 60 trials per participant. The modality of experimental sessions was counterbalanced between participants. Each experimental session was divided into two 30-trial blocks containing 15 related trials and 15 unrelated trials. Participants only had to recognise one type of transformation per block, therefore one block contained all the related inverse trials, and the other block contained all related retrograde trials. The order of blocks was counterbalanced between participants, and the presentation order of trials within each block was also randomised.

On arrival participants completed a brief questionnaire collecting demographic information pertaining to age, gender, handedness, potential hearing problems, and musical experience. Participants were then seated in front of a PC monitor and taken through a series of instructions by the experimenter. Before each block, participants were introduced to the relevant transformation and presented with examples. They then took part in 6 untimed example trials with the experimenter present. Participants were instructed to focus on both patterns and decide whether the target was 'related' (i.e. whether it was an inverse or retrograde transformation) or 'unrelated' to the standard. They indicated their decision by pressing one of two buttons on a response box using their index and middle fingers of their dominant hand. In addition, they were instructed to respond as quickly as possible whilst maintaining accuracy. In half of the experimental sessions 'related' responses were allocated to the left button, and in the other half 'related' responses were allocated to the right button. When the participant was ready to start the experimental trials, the experimenter left the room. First, participants took part in 6 timed practice trials. Feedback was provided for responses to these timed practice trials, but not for responses to the experimental trials. When the participant had finished the first block, the experimenter returned to provide further instructions specific to the upcoming block, which involved a different transformation. The entire experimental session took approximately 25 minutes to complete.

5.2.2 Results

Data from two participants were excluded from analysis because they failed to perform above chance levels (overall error rate 50% or greater). Paired-samples t-tests were carried out to examine any effects of block order (first block, second block) on overall PE (arcsine-transformed) and RT (log-transformed). There was a significant effect of block order on PE, $t(55) = 0.01$, $p = .028$, one-tailed, with more errors being made in the first block (M = 27.14, SD = 14.28) compared to the second block (M = 23.93, SD = 13.15). However, the effect of block order on RT failed to reach significance, $t(55) = 0.44$, $p = .331$, one-tailed.

5.2.2.1 ***Error data***

5.2.2.1.1 All trials

Overall PE was 25.54, which was approximately 5% less than was observed in Experiment 2 (unimodal trials with horizontally presented visual stimuli). This may have been due to the greater proportion of skilled participants in the sample (in the present experiment 48% of participants had received a mean of 4.38 years music training, whereas in Experiment 2 only 20% of participants had received a mean of 3.48 years music training) – though an exploratory analysis on music training failed to reveal any significant effects. Mean results for responses to retrograde, inverse and unrelated targets in both modality conditions are displayed in Figure 5.2.

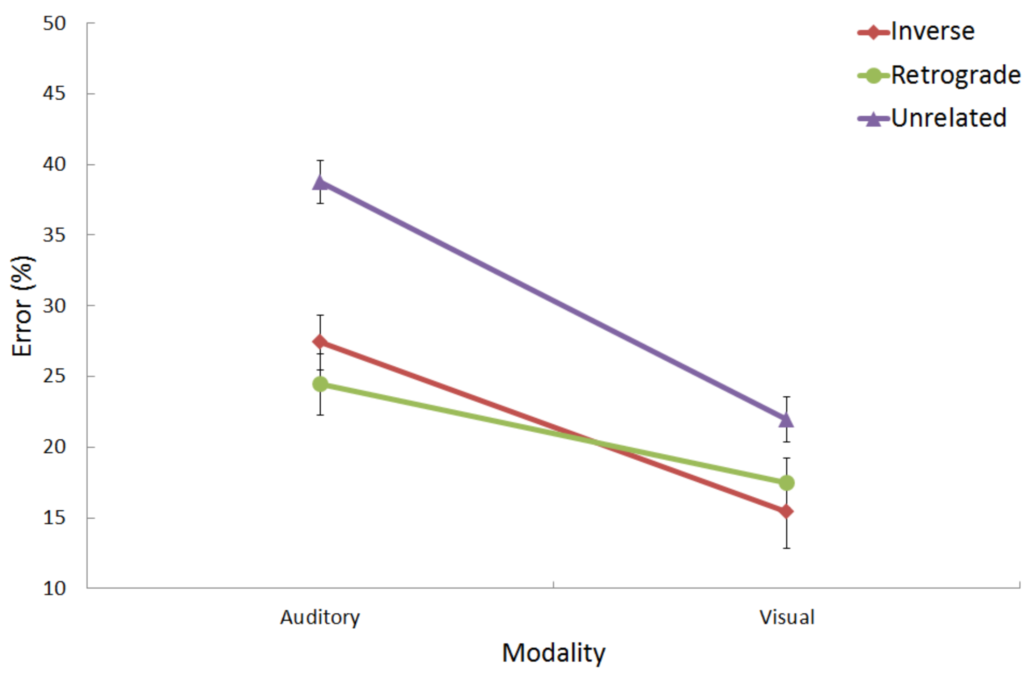


Figure 5.2. Experiment 4: Mean PE in target conditions, plotted as a function of modality. Error bars indicate the standard error of the mean (SEM ± 1).

A preliminary 2 x 2 mixed ANOVA examined the effects of the within-subjects factor relatedness (related, unrelated) and the between-subjects factor modality (auditory, visual) on arcsine-transformed PE. The main effect of relatedness was highly significant, $F(1,54) = 16.92$, $MSE = .02$, $p < .001$, $\eta_p^2 = .24$, with more errors being made when responding to unrelated ($M = 30.06$, $SD = 17.74$) than related targets ($M = 21.01$, $SD = 11.72$). The main effect of modality was also highly significant, $F(1,54) = 19.27$, $MSE = .04$, $p < .001$, $\eta_p^2 = .26$, with more errors being made in the auditory ($M = 32.36$, $SD = 15.60$) than the visual condition ($M = 19.20$, $SD = 16.36$). The interaction between relatedness and modality failed to reach significance, $F(1,54) = 2.02$, $MSE = .04$, $p = .161$, $\eta_p^2 = .04$.

5.2.2.1.2 Related trials only

The above significant main effect of relatedness permitted the performance of further analysis to examine the 1½-D hypothesis. A 2 x 2 mixed ANOVA was run to examine the between-subjects effect of modality (auditory, visual) and the within-subjects effect of transformation (inverse, retrograde) when participants were responding to related targets only. The analysis revealed a significant main effect of modality, $F(1,54) = 9.63$, $MSE = .06$, $p = .003$, $\eta_p^2 = .15$. Participants made more errors in the auditory ($M = 25.93$, $SD = 11.82$) than the visual condition ($M = 16.44$, $SD = 9.76$). The 1½-D hypothesis predicted that inverse transformations would be recognised more successfully than retrograde transformations. However, the main effect of transformation failed to reach significance, $F(1,54) = 0.22$, $MSE = .02$, $p = .642$, $\eta_p^2 < .01$. The interaction between transformation and modality also failed to reach significance, $F(1,54) = 0.47$, $MSE = .02$, $p = .498$, $\eta_p^2 = .01$.

Despite the absence of a significant main effect of transformation or a significant interaction, pairwise comparisons were carried out to examine the simple effects of transformation, and hence the 1½-D hypothesis, more closely. The mean differences failed to reach significance in either modality condition (auditory [$MD = 2.96$, $SE = 2.93$; $p = .428$]; visual [$MD = 2.07$, $SE = 2.82$; $p = .878$]). No further analysis was carried out.

5.2.2.2 ***RT data***

5.2.2.2.1 All trials

Overall mean RT was 841.52ms, which was slightly faster (by 46ms) than the overall response time observed in Experiment 2. This is in agreement with the error data – when compared to Experiment 2 participants made fewer errors and were quicker to identify targets in Experiment 4. Figure 5.3 shows the mean RT across all experimental conditions.

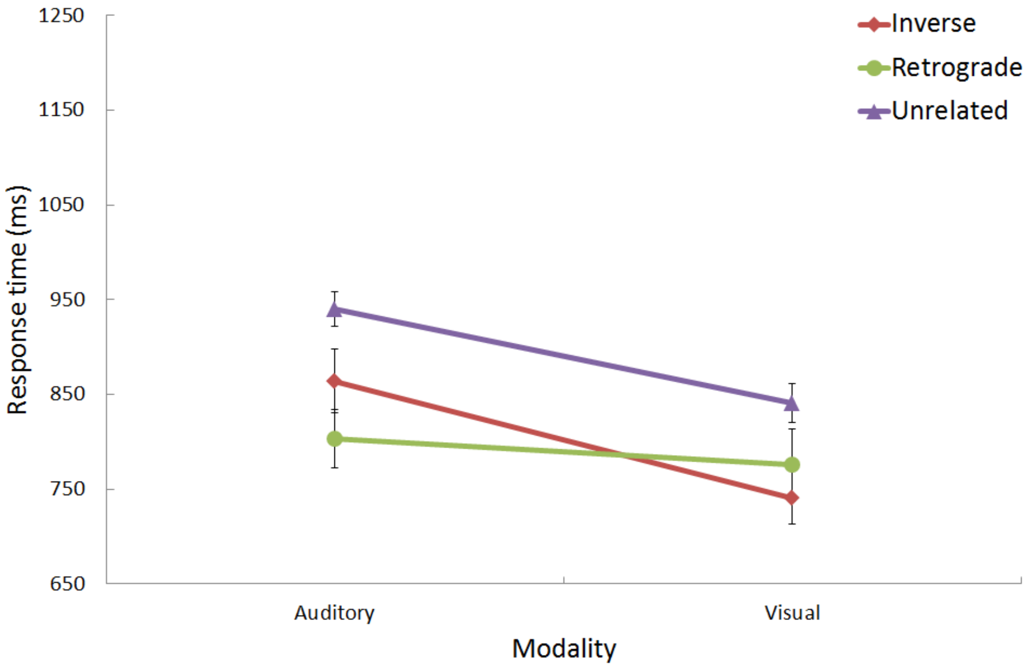


Figure 5.3. Experiment 4: Mean RT in target conditions, plotted as a function of modality. Error bars indicate the standard error of the mean (SEM ± 1).

The 2 x 2 mixed ANOVA (run on log-transformed RT data), with the within-subjects factor relatedness and the between-subjects factor modality, revealed a highly significant main effect of relatedness, $F(1,54) = 13.44$, $MSE = .03$, $p = .001$, $\eta_p^2 = .20$, with slower responses to unrelated ($M = 888.56$, $SD = 468.67$) compared to related targets ($M = 794.49$, $SD = 401.07$). Mean RTs were slower in the auditory condition ($M = 886.85$, $SD = 435.62$) than the visual condition ($M = 799.33$, $SD = 416.65$), but the main effect of modality failed to reach significance, $F(1,54) = 0.50$, $MSE = .64$, $p = .481$, $\eta_p^2 = .01$. The interaction between relatedness and modality also failed to reach significance, $F(1,54) = 0.75$, $MSE = .03$, $p = .389$, $\eta_p^2 = .01$.

5.2.2.2.2 Related trials only

A further 2 x 2 mixed ANOVA was carried out in order to examine the within-subjects effect of transformation and the between-subjects effect of modality on RT for related targets only. Once again, mean RT was slower in the auditory condition ($M = 833.54$, $SD = 397.18$) than in the visual condition ($M = 758.14$, $SD = 408.23$), but the main effect of modality failed to reach significance, $F(1,54) = 0.73$, $MSE = .72$, $p = .397$, $\eta_p^2 = .01$. The main effect of transformation was non-significant, $F(1,54) < 0.01$, $MSE = .12$, $p = .950$, $\eta_p^2 < .01$, and thus the RT data failed to support the 1½-D hypothesis. The interaction between transformation and modality also failed to reach significance, $F(1,54) = 0.60$, $MSE = .12$, $p = .441$, $\eta_p^2 = .01$.

Pairwise comparisons were run to examine the simple effects of transformation more closely. The mean differences were non-significant in both modality conditions (auditory [$MD = 60.90$, $SE = 54.42$; $p = .562$]; visual [$MD = 34.10$, $SE = 52.51$; $p = .609$]). This pattern of results mirrors that revealed by the analysis on PE data. No further analysis was carried out.

5.2.2.3 ***Signal detection analysis***

As was done in the experiments reported in the previous chapter, further analysis was carried out using signal detection theory.

5.2.2.3.1 Sensitivity to the signal

The analysis on *d′* was in agreement with the analysis on PE data. Overall *d′* was 1.35, which was higher than overall *d′* in Experiment 2 – targets were more detectable in the present experiment. The mean results in each condition are displayed in Figure 5.4. The 2 x 2 ANOVA revealed a highly significant main effect of modality, $F(1,54) = 16.60$, $MSE = 1.11$, $p < .001$, $\eta_p^2 = .24$, with detectability being better in the visual condition ($M = 1.74$, $SE = .14$) than in the auditory condition ($M = 0.93$, $SE = .14$). The main effect of transformation failed to reach significance, $F(1,54) = 0.41$, $MSE = 0.33$, $p = .525$, $\eta_p^2 = .01$, as did the interaction between transformation and modality, $F(1,54) = 0.41$, $MSE = 0.33$, $p = .526$, $\eta_p^2 = .01$.

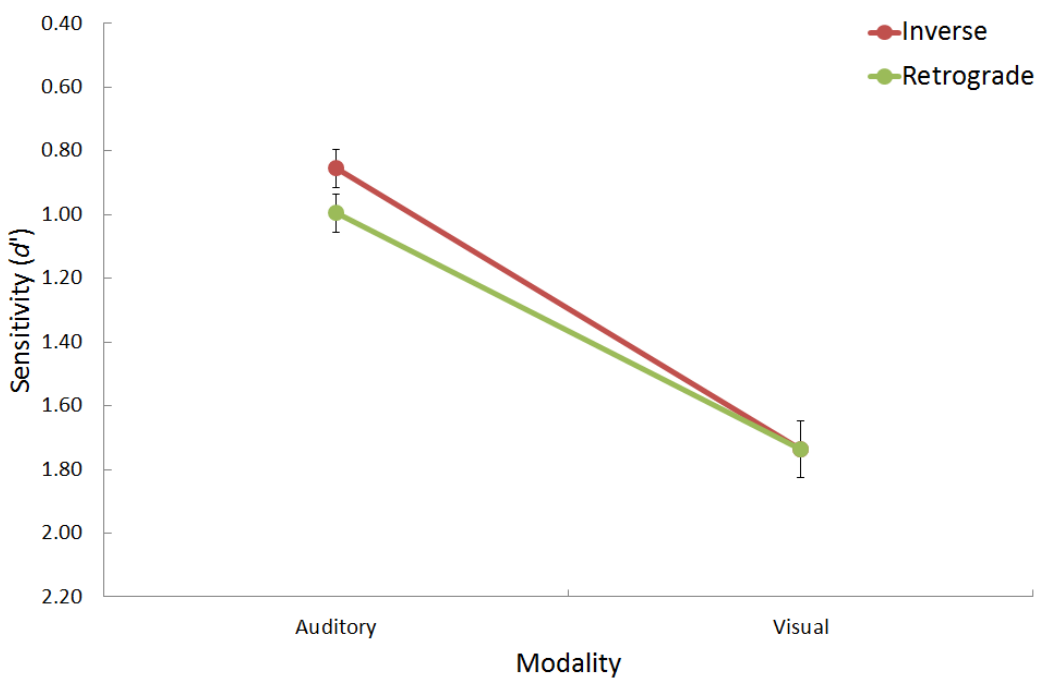


Figure 5.4. Experiment 4: Mean *d′* in transformation conditions, plotted as a function of modality. The scale along the y-axis has been inverted so that results can be more easily compared with figures displaying PE data. Error bars indicate the standard error of the mean (SEM ± 1).

Pairwise comparisons were run to examine the simple effects of transformation in different modality conditions, but failed to reveal any significant differences (auditory [MD = 0.14, SE = 0.16; p = .378]; visual [MD = 0.00, SE = 0.15; p = .999]). No further analysis was carried out.

5.2.2.3.2 Response bias

Mean *c* was -0.10. Though participants were still biased towards ‘same’ responses, they were less liberal than in Experiment 2. The 2 x 2 ANOVA failed to reveal any significant effects.

5.2.2.4 *Music training analysis*

An exploratory analysis of the effects of music training was carried out by running correlations and repeating the ANOVAs run on PE and RT data from related trials only, and on d' and c data, with music training included as a between-subjects factor (see Appendix IV for the full results – only significant findings will be reported here). 26 participants self-reported some previous music training and were allocated to the some training group. 30 participants self-reported no previous music training and were allocated to the no training group.

Firstly, a Pearson's correlation analysis revealed a number of significant correlations between amount of music training and performance, but only when responding to retrograde transformations in the visual condition. Increasing music training was associated with increasing PE, $r = .85$, $p < .001$, increasing RT, $r = .59$, $p = .035$, and decreasing d', $r = -.62$, $p = .001$. So, as music training increased performance got worse. These results were contrary to the processing advantage that was expected to be associated with music training. A significant positive correlation was also found between music training and c data in the same condition, $r = .79$, $p = .001$, indicating that increasing training was associated with decreasingly liberal responses. The same association had been found in Experiments 2 and 3, but for responses to inverse transformations in visual and VA conditions.

Secondly, a series of 2 x 2 x 2 mixed ANOVAs, with the within-subjects factors modality (auditory, visual) and transformation (inverse, retrograde) and the between-subjects factor music training (some training, no training), were run on PE, RT, d' and c data. The results of interest were the main effect of music training and the interactions between music training and other factors. Any significant interactions were followed up with pairwise comparisons exploring the simple effects of music training.

The analysis on PE data failed to reveal any significant results. The analysis on RT data revealed a significant three-way interaction between transformation, modality and music training, $F(1,52) = 4.73$, $MSE = .12$, $p = .034$, $\eta_p^2 = .08$. Pairwise comparisons were performed to examine the simple effects of music training in the different conditions, but all comparisons failed to reach significance. The analysis on *d′* data revealed a significant interaction between modality and music training, $F(1,52) = 4.87$, $MSE = 1.04$, $p = .032$, $\eta_p^2 = .09$. Pairwise comparisons were run to examine the simple effects of music training in both modality conditions. In the visual condition, the mean difference between music training groups reached significance ($MD = 0.59$, $SE = 0.27$; $p = .034$), with detection being better in the 'some training' group ($M = 2.06$, $SD = 0.74$) than in the 'no training' group ($M = 1.47$, $SD = 0.89$). In the auditory condition, the mean difference between music training groups failed to reach significance.

The analysis on *c* revealed a significant three-way interaction between transformation, modality and music training, $F(1,52) = 5.02$, $MSE = .05$, $p = .029$, $\eta_p^2 = .09$. Pairwise comparisons examining the simple effects of music training in all conditions revealed a significant mean difference between music training groups when detecting inverse transformations of visual targets ($MD = 0.30$, $SE = 0.13$; $p = .022$). When detecting inverse transformations of visual targets participants with some training were conservative in their responses ($M = 0.13$, $SD = 0.27$) and participants with no training were liberal ($M = -0.17$, $SD = 0.49$). In other words, when detecting inverse transformations of visual targets, participants with some music training were more likely to indicate that targets were unrelated, but participants with no training were more likely to indicate that targets were related. In all other conditions participants in both music training groups adopted a similar bias strategy.

5.2.3 Discussion

The aim of Experiment 4 was to examine the processing of inverse and retrograde transformations when auditory and visual stimuli corresponded to equivalent 1½-D supramodal pattern spaces. In both modality conditions, the 1½-D hypothesis was tested, which predicted that inverse transformations would be recognised more successfully than retrograde transformations.

The results from Experiment 4 failed to support the 1½-D hypothesis, as no significant effects of transformation were revealed by the analysis. The prediction made by the 1½-D hypothesis was based on the assumption that recognition would be based on the processing of structural information, and that inversions of ordinal relations on the temporal dimension are harder to process than inversions on the scalar dimension. The absence of an effect of transformation in the present experiment challenges this assumption, and might indicate that inversions on the temporal dimension are actually no harder to process than inversions on the scalar dimension.

It is possible that the observed results were influenced by the availability of non-structural information, which facilitated the recognition of retrograde transformations. As was highlighted and discussed in Chapter 4, when auditory targets are presented in the same pitch space as standards, their tones preserve the pitches of standard tones when a retrograde transformation has been applied, albeit in reverse order. In the present experiment, visual stimuli comprised sequences of objects that were all presented centrally, and as a result target objects under retrograde transformation preserved the spatial positions of the standard objects. In short, non-structural information could have facilitated the recognition of retrograde transformations in both modality conditions of the present experiment.

To eliminate the possibility that recognition performance in unimodal trials is contaminated by the processing of non-structural information in the retrograde transformation condition, the experiment would need to be repeated using target stimuli that are transposed to different pitch registers or different spatial positions. Another way of eliminating the potentially contaminating effects of non-structural information would be to present patterns in cross-modal trials. For example, when patterns were presented cross-modally in Experiment 3, some support was found for the 1½-D hypothesis in the AV condition, which was attributed to the processing of auditory standards that corresponded to a 1½-D supramodal pattern space.

The absence of a transformation effect on the auditory condition of Experiment 4 contrasted with the finding that retrograde transformations were recognised more successfully than inverse transformations in the auditory condition of Experiment 2. If the availability of non-structural information obscured what would otherwise have been a transformation effect indicating a processing advantage for inverse transformations, why would the facilitation effect have been greater in Experiment 2 (reversing the predicted transformation effect) than in Experiment 4 (merely cancelling out the predicted transformation effect)? It is difficult to find a sufficient explanation. The only difference between the auditory condition in Experiment 2 and 4 was that in the latter participants only took part in one modality condition. It is possible that in Experiment 2 responses were influenced by the explicit analogy between auditory and visual stimuli, and were encouraged to 'visualise' auditory stimuli or 'audiate' visual stimuli.

This interpretation logically leads to a reassessment of the retrograde advantage observed in the auditory condition of Experiment 2. It is possible that the within-subjects modality condition encouraged participants to 'visualise' auditory patterns, in line with the way in which they were presented in that experiment (horizontally on the screen). They then could have used the same strategy to identify retrograde targets that was proposed to explain results in the VA condition of Experiment 3. Namely, they could have retraced a 2-D representation of the standard, which would not have required mental transformation of the pattern. However, it must be conceded that these alternative interpretations are highly speculative, and no meaningful conclusions can be made without controlling for the potential influence of non-structural information.

It should be noted that the absence of a transformation effect in the visual condition of the present experiment replicated the results found in the visual condition of Experiment 2. Therefore, despite the change made to the way in which visual stimuli were presented, transformations of patterns that corresponded to different supramodal pattern spaces (1½-D and 2½-D, respectively) were apparently processed similarly. Once again, it is difficult to find a sufficient explanation for these results. The absence of a transformation effect in Experiment 4 cannot be explained by the 2½-D hypothesis that was used to interpret the results in Experiment 2. Neither can the result in the present experiment be explained by other theories of visual perception, such as the transformational approach (Hahn, Chater, & Richardson, 2003; Hahn, 2014; Palmer, 1983). The transformational approach applies to the perception of pattern regularities in 2-D visual images, and can therefore not be easily extended to the perception of the visual stimuli used in Experiment 4, which presented sequential objects on a single vertical spatial dimension.

Moving on from the effects of transformation and considering the general effects of modality, analysis on PE and *d′* data demonstrated a clear performance advantage in the visual condition. This is in agreement with the results of Experiment 2, and further supports the view that structural information may be abstracted more efficiently from visual stimuli (Balch & Muscatelli, 1986). Interestingly, the analysis on RT data failed to replicate the visual advantage observed in Experiment 2, though there are no clear explanations for why this should be the case. Another interesting finding was revealed by the analysis on music training, which demonstrated that participants with some music training were better at detecting transformations of visual patterns. Surprisingly, this detection advantage did not extend to auditory patterns. As performance in visuo-spatial tasks has previously been linked to music training, with 'musicians' performing better than 'non-musicians' (Pietsch & Jansen, 2012), it is not clear how to interpret this finding – if participants performed better in the visual condition because of previous music training, why did they not also perform better in the auditory condition? Furthermore, this pattern of results runs contrary to that found in Experiment 2 and other research, which has demonstrated that participants with some training are better at processing auditory patterns (Halpern, Bartlett, & Dowling, 1998; Trainor, Desjardins, & Rockel, 1999).

In conclusion, the results of Experiment 4 failed to support the 1½-D hypothesis. Although these results challenge the assumptions of the SPS framework, it is possible that a processing advantage for inverse transformations was obscured by the facilitation effects of redundant non-structural information, when recognising retrograde transformations. Therefore, no meaningful conclusions can be made without first eliminating the potential influence of non-structural information. This may be achieved by transposing targets in unimodal trials, or by presenting auditory and visual patterns in cross-modal trials. These possibilities were tested in Experiments 5 and 6.

5.3 Experiment 5: Cross-modal trials

Experiment 5 was a replication of Experiment 3, using centrally presented visual stimuli that did not map the timing of events onto the horizontal dimension. Auditory and visual stimuli were presented in cross-modal trials: auditory standards were followed by visual targets (AV condition) or visual standards were followed by auditory targets (VA condition). As both auditory and visual stimuli could be represented in a 1½-D supramodal pattern space, the 1½-D hypothesis was tested in both modality conditions. The hypothesis predicted that recognition would be better for targets under inverse transformation.

5.3.1 Methods

5.3.1.1 *Participants*

42 students from the University of Roehampton took part in Experiment 5 (all female; mean age = 25.71 years, *SD* = 12.19). All had normal hearing and

normal or corrected-to-normal vision. Five participants reported they were left-handed, one ambidextrous, and the remainder were right-handed. 25 participants (60%) reported some level of music training (mean = 4.37 years). They all received course credit for their participation.

5.3.1.2 *Stimuli*

The stimuli used in Experiment 3 were identical in all aspects to those presented in Experiment 4, except with respect to the modality of standard and target patterns. In Experiment 4 the trials were unimodal, but in Experiment 3 the trials were cross-modal.

5.3.1.3 *Design and procedure*

The design and procedure were identical to those of Experiment 3. In Experiment 5 each trial was cross-modal (auditory-visual [AV], visual-auditory [VA]). Trials in the AV condition comprised an auditory standard followed by a visual target, and trials in the VA condition comprised a visual standard followed by an auditory target. As both auditory and visual stimuli were presented in these cross-modal trials, there was little point in making modality a between-subjects factor, as it had been in Experiment 4. Therefore, the factor modality was once again within-subjects.

5.3.2 Results

Data from four participants were excluded from analysis because they failed to perform above chance levels (overall error rate 50% or greater). Paired-samples *t*-tests were carried out to examine any effects of block order. The effect of block order on PE (arcsine-transformed) approached significance, $t(37) = 1.59$, $p = .060$, one-tailed, with higher PE in the first block ($M = 28.25$, $SD = 12.19$) than in the second block ($M = 25.26$, $SD = 14.27$). Mean RT in the first block ($M = 1003.99$, $SD = 320.74$) was marginally slower than in the second block ($M = 990.82$, $SD = 410.47$), but the effect failed to reach significance, $t(37) = 0.89$, $p = .190$, one-tailed.

5.3.2.1 ***Error data***

5.3.2.1.1 All trials

Overall PE was 26.75. This was approximately 5% higher than was recorded for responses in Experiment 3 (cross-modal trials with horizontally presented visual stimuli). Mean results for responses to retrograde, inverse and unrelated targets in both modality conditions are displayed in Figure 5.5.

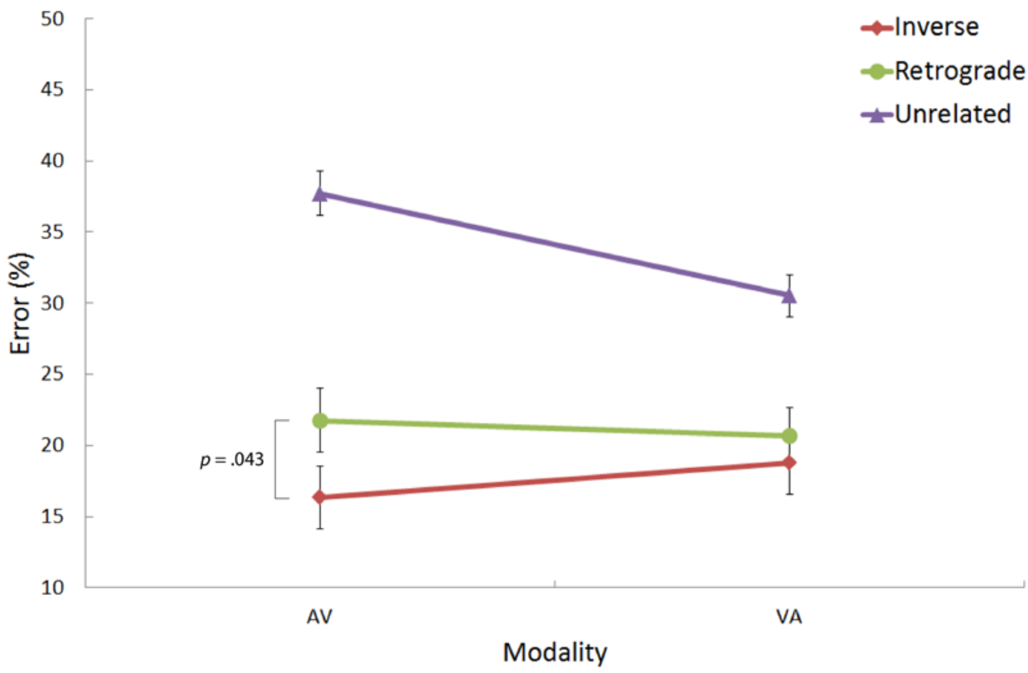


Figure 5.5. Experiment 5: Mean PE in target conditions, plotted as a function of modality. Significance values for simple effects were obtained from the analysis on arcsine-transformed data. Error bars indicate the standard error of the mean (SEM ± 1).

A 2 x 2 ANOVA was carried out in order to examine the within-subjects effects of relatedness and modality. As in previous experiments, there was a highly significant main effect of relatedness, $F(1,37) = 40.58$, $MSE = .03$, $p < .001$, $\eta_p^2 = .52$. Participants made more errors when responding to unrelated targets ($M = 34.12$, $SD = 15.91$) compared to related targets ($M = 19.39$, $SD = 9.99$). Although participants made slightly more errors overall in the AV condition ($M = 28.38$, $SD = 12.37$) compared to the VA condition ($M = 25.13$, $SD = 12.26$), the main effect of modality failed to reach significance, $F(1,37) = 2.17$, $MSE = .02$, $p = .150$, $\eta_p^2 = .06$. There was a significant interaction between relatedness and modality, $F(1,37) = 10.15$, $MSE = .01$, $p = .003$, $\eta_p^2 = .22$. The interaction was due to a greater effect of relatedness in the AV compared to the VA condition. Pairwise comparisons confirmed the significant effect of relatedness in both modality conditions – the mean difference between relatedness conditions was highly significant in both the AV (*MD* = 18.68, *SE* = 2.61; $p < .001$) and the VA condition (*MD* = 10.79, *SE* = 2.37; $p < .001$).

5.3.2.1.2 Related trials only

A further 2 x 2 ANOVA was carried out in order to examine the effects of transformation and modality on PE for related targets only. The main effect of modality was non-significant, $F(1,37) = 0.20$, $MSE = .03$, $p = .655$, $\eta_p^2 = .01$. The 1½-D hypothesis predicted that lower PE would be observed for inverse transformations. However, the main effect of transformation failed to reach significance, $F(1,37) = 1.72$, $MSE = .05$, $p = .198$, $\eta_p^2 = .04$. The interaction between transformation and modality was also non-significant, $F(1,37) = 2.29$, $MSE = .04$, $p = .138$, $\eta_p^2 = .06$.

Despite the absence of a significant main effect of transformation, or a significant interaction, pairwise comparisons were run to examine the simple effects of transformation in both modality conditions. This was done because the 1½-D hypothesis made specific predictions about the effect of transformation in both modality conditions. As can be seen in Figure 5.5, mean PE was lower for inverse transformations in both modality conditions. A significant mean difference was found in the AV condition (*MD* = 5.44, *SE* = 3.37; $p = .043$). This result provided some support for the 1½-D hypothesis. However, in the VA condition the 1½-D hypothesis was unsupported, as the mean difference failed to reach significance (*MD* = 1.93, *SE* = 3.24; $p = .971$).

5.3.2.2 ***RT data***

5.3.2.2.1 All trials

Overall mean RT was 997.17ms, which was approximately 68ms faster than overall responses to cross-modal trials in Experiment 3. Thus, although more errors were made in Experiment 5 compared to Experiment 3, targets were identified more quickly. Figure 5.6 shows the mean RT across all experimental conditions.

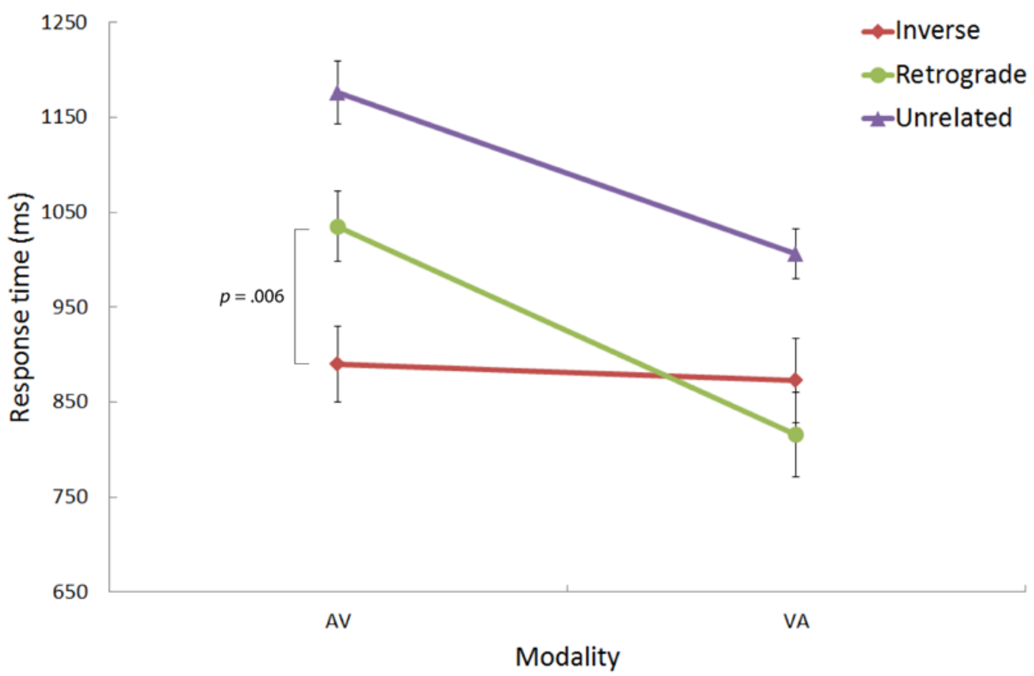


Figure 5.6. Experiment 5: Mean RT in target conditions, plotted as a function of modality. Significance values for simple effects were obtained from the analysis on log-transformed data. Error bars indicate the standard error of the mean (SEM ± 1).

The initial 2 x 2 ANOVA revealed a significant main effect of relatedness, $F(1,37) = 35.78$, $MSE = .04$, $p < .001$, $\eta_p^2 = .49$. Participants were slower to identify unrelated targets ($M = 1091.03$, $SD = 376.06$) than related targets ($M = 925.17$, $SD = 362.65$). The main effect of modality was significant, $F(1,37) =$

15.69, *MSE* = .08, p = .001, η_p^2 = .30, with slower responses recorded in the AV (*M* = 1069.18, *SD* 365.26) than in the VA condition (*M* = 925.17, *SD* = 362.65). The interaction between relatedness and modality failed to reach significance, $F(1,37) = 0.32$, *MSE* = .03, p = .576, η_p^2 = .01.

5.3.2.2.2 Related trials only

A further 2 x 2 ANOVA was carried out in order to examine the effects of modality and transformation on RT for related targets only. The main effect of modality was significant, $F(1,37) = 10.31$, *MSE* = .10, p = .003, η_p^2 = .22, with slower responses in the AV (*M* = 962.32, *SD* = 345.10) than in the VA condition (*M* = 844.31, *SD* = 372.76). Once again, the main effect of transformation failed to reach significance, $F(1,37) = 0.99$, *MSE* = .09, p = .326, η_p^2 = .03. However, there was a significant interaction between transformation and modality, $F(1,37) = 10.12$, *MSE* = .06, p = .003, η_p^2 = .22.

Visual inspection of Figure 5.6 shows that mean RT was faster for inverse tranformations in the AV condition, but slower for inverse transformations in the VA condition. Pairwise comparisons were run to examine the simple effects of transformation in modality conditions. In the AV condition the mean difference was significant (*MD* = 145.51, *SE* = 55.96; p = .006), supporting the 1½-D hypothesis. In the VA condition the mean difference failed to reach significance (*MD* = 57.21, *SE* = 68.33; p = .239). The pattern of results revealed by the analysis on RT data is similar to that revealed by the analysis on PE data (despite the interaction between modality and transformation not being significant). Taken together, they both provide some support for the 1½-D hypothesis in the AV condition, but fail to support the 1½-D hypothesis in the VA condition.

Pairwise comparisons were also run to examine the simple effects of modality. There was a highly significant mean difference in the retrograde condition (*MD* = 219.36, *SE* = 59.60; $p < .001$), with responses being faster in the VA condition, but the mean difference in the inverse condition failed to reach significance (*MD* = 16.64, *SE* = 57.61; $p = .520$).

5.3.2.3 ***Signal detection analysis***

Further analysis was carried out using signal detection theory.

5.3.2.3.1 Sensitivity to the signal

The analysis on d' was largely in agreement with the analysis on PE data. Mean d' was 1.34, which was slightly lower than in Experiment 3 – targets were less detectable in the present experiment. The mean results in each condition are displayed in Figure 5.7. The two-way ANOVA revealed that the main effect of modality was approaching significance, $F(1,37) = 3.20$, $MSE = .39$, $p = .082$, $\eta_p^2 = .08$, with targets being more detectable in the VA condition ($M = 1.43$, $SE = .13$) than in the AV condition ($M = 1.25$, $SE = .14$). The main effect of transformation failed to reach significance, $F(1,37) = 0.69$, $MSE = .77$, $p = .414$, $\eta_p^2 = .02$. However, the interaction between transformation and modality was approaching significance, $F(1,37) = 2.94$, $MSE = .52$, $p = .095$, $\eta_p^2 = .07$.

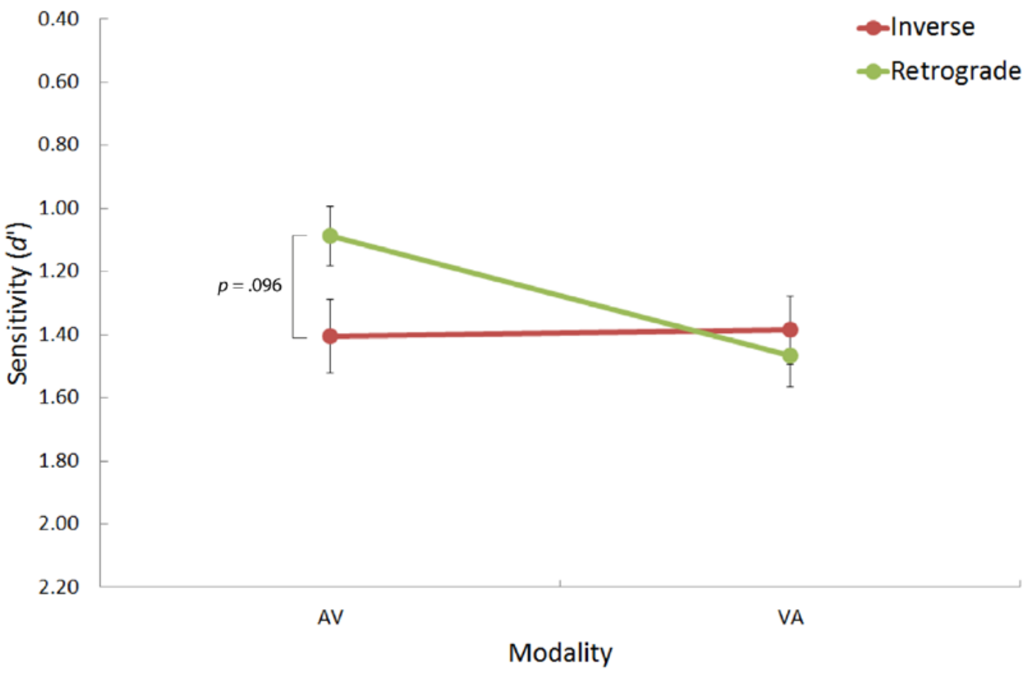


Figure 5.7. Experiment 5: Mean *d′* in transformation conditions, plotted as a function of modality (NB simple effects of modality are not displayed). The scale along the y-axis has been inverted so that results can be more easily compared with figures displaying PE data. Error bars indicate the standard error of the mean (SEM ± 1).

Visual inspection of Figure 5.7 shows that the mean detection rates for inverse transformations were higher in the AV condition (indicating greater sensitivity), but lower in the VA condition. Although the transformation effect in the AV condition was in the predicted direction, pairwise comparisons revealed that the mean difference did not reach formal significance ($MD = 0.32$, $SE = .19$; $p = .096$), failing to support the 1½-D hypothesis. The mean difference in the VA condition also failed to reach significance ($MD = 0.08$, $SE = .18$; $p = .652$).

5.3.2.3.2 Response bias

Mean *c* was -0.20, which was markedly more liberal than in Experiment 3 – participants in the present experiment were more biased towards ‘same’ responses. The 2 x 2 ANOVA revealed a significant main effect of modality on response bias, $F(1,37) = 6.38$, $MSE = .07$, $p = .016$, $\eta_p^2 = .15$, with more liberal responses being made in the AV condition ($M = -0.26$, $SE = .04$) compared to the VA condition ($M = -0.15$, $SE = .04$). This contrasted with the results of Experiment 3, which found no significant main effect of modality. The main effect of transformation and the interaction between transformation and modality both failed to reach significance.

5.3.2.4 ***Music training analysis***

To explore the effects of music training on performance, 22 participants were allocated to the some training group, and 16 participants were allocated to the no training group. Firstly, a Pearson's correlation analysis was run on the amount of training reported by participants in the 'some training' group, and performance in the experimental conditions. A number of significant correlations were revealed. There was a significant negative correlation with PE data, when responding to retrograde transformations in the VA condition, $r = -.46$, $p = .027$ – increasing levels of music training was associated with decreasing error rates. There was a significant positive correlation with d' data in the same condition, $r = .56$, $p = .005$, indicating that increasing levels of music training was associated with greater detection. A further significant positive correlation with d' data was revealed when responding to inverse transformations in the AV condition, $r = .44$, $p = .037$. Thus, increasing levels of music training was associated with increasing detection scores in different transformation and modality conditions. A final significant correlation with c data was observed, also when responding to inverse transformations in the AV condition, $r = .43$, $p = .039$. Thus, in this condition increasing levels of music training were associated with decreasingly liberal responses.

Secondly, a series of 2 x 2 x 2 mixed ANOVAs, with the within-subjects factors modality (AV, VA) and transformation (inverse, retrograde) and the between-subjects factor music training (some training, no training), were run on PE, RT, *d′* and *c* data. The analysis on PE data revealed a significant three-way interaction between modality, transformation, and music training, $F(1,36) = 5.55$, $MSE = .04$, $p = .024$, $\eta_p^2 = .13$. Pairwise comparisons were run to examine the simple effects of music training in all conditions, but failed to reveal any significant results. The analysis on RT data failed to reveal any significant results. The analysis on *d′* data revealed a significant main effect of music training, $F(1,36) = 7.75$, $MSE = 1.94$, $p = .008$, $\eta_p^2 = .18$. Participants in the some training group ($M = 1.61$, $SD = 0.76$) were better at detecting transformed targets than participants in the no training group ($M = 0.97$, $SD = 0.60$). Finally, the analysis on *c* data also revealed a significant main effect of music training, $F(1,36) = 6.80$, $MSE = 0.15$ $p = .013$, $\eta_p^2 = .16$. Participants in both music training groups were liberal when responding to targets (they were biased towards indicating that targets were related), but participants in the no training group were more liberal than participants in the some training group (no training: $M = -0.30$, $SD = 0.13$; some training: $M = -0.14$, $SD = 0.23$).

5.3.3 Discussion

The results of Experiment 5 provided some support for the 1½-D hypothesis. In the AV condition, performance was consistently better for inverse transformations (although the analysis on *d′* data failed to reach formal significance). However, in the VA condition no effect of transformation was observed – inverse and retrograde transformations were recognised equally well. This result failed to support the 1½-D hypothesis.

According to the SPS framework, the stimuli presented in the experiment can be represented in a 1½-D supramodal pattern space. Inverse transformation of patterns represented in a 1½-D space requires an inversion of ordinal relations on a scalar dimension, whereas retrograde transformation requires an inversion of ordinal relations on a temporal dimension, which is harder to process due to the temporal dimension's inherent directionality. This would explain the recognition advantage for targets under inverse transformation in the AV condition. The finding is in agreement with the results reported in Experiment 3, which also demonstrated a performance advantage for inverse transformations in the AV condition (though the effect only reached formal significance in the analysis on PE data).

Based on participant comments and the design of the experiment, it has been assumed that the recognition of targets in the experiments reported in this thesis is based on the mental transformation of structural representations of the standard pattern. When interpreted within the SPS framework, the pattern of results in the AV conditions of Experiments 3 and 5 support this assumption. If processing had been based on structural abstractions from the target, then contrasting effects of transformation would have been observed in different experiments – target patterns (visual stimuli) could be represented in a 2½-D (Experiment 3) or a 1½-D space (Experiment 5). Instead, the results in the AV conditions of both experiments supported the 1½-D hypothesis, suggesting that recognition was based on the processing of standard patterns (auditory stimuli) represented in 1½-D supramodal pattern space.

Returning to the present experiment, it is not clear why the 1½-D hypothesis was supported in the AV condition but not in the VA condition. Whether recognition in the VA condition was based on the processing of structural representations of standard or target patterns, the 1½-D hypothesis predicted that inverse transformations should have been recognised more effectively, as both auditory and visual stimuli could be represented in a 1½-D supramodal pattern space. Although mean error was lower for inverse transformations, mean RT was higher and mean d' was lower (indicating better detection for retrograde transformations). Regardless, all analyses were consistent in demonstrating no significant effect of transformation on performance.

The absence of a transformation effect is more in line with the predictions made by the 2½-D hypothesis and the facilitation effect of sensory-specific, non-structural information. Yet, the result cannot be explained by the 2½-D hypothesis because visual stimuli did not correspond to a 2½-D representation. Furthermore, the result cannot be explained by the effect of non-structural information because this was not available in cross-modal trials. An alternative explanation can be sought in possible 'visualisation' strategies that have been discussed in previous experiments. For example, in the VA condition of Experiment 3, when visual stimuli corresponded to 2½-D representations, retrograde transformations were recognised more effectively than inverse transformations. When discussing the results of Experiment 3 (Section 4.3.3), it was suggested that the performance advantage for retrograde transformations may have been due to participants being able to employ an alternative visualisation strategy whereby 2-D representations of visual standards were held in memory and retraced as auditory targets unfolded. Thus, by presenting visual stimuli horizontally, auditory targets under retrograde transformation could have been identified without having to mentally transform the visual standard.

Following the same line of thought that was used to interpret the findings in Experiment 3, the failure to find an advantage for inverse transformations in the VA condition of the present experiment could be attributed to a different type of visualisation strategy. All auditory stimuli used in the experimental trials consisted of the same three tones, and all visual stimuli consisted of objects at the same three spatial positions. It is not unreasonable to assume that participants were able to learn to associate the pitches of tones with the spatial positions of visual objects. Should this have been the case, then auditory targets could have been visualised and recognition could have been based on additional sensory-specific, non-structural information (i.e. visualised targets would preserve the spatial positions of standards, albeit in reverse order). In this case, the results would be explained by the availability of additional redundant information, which facilitated the perception of retrograde transformations, cancelling out the transformation effect predicted by the 1½-D hypothesis.

This alternative explanation illustrates how sensory-specific, non-structural information could apply to cross-modal trials – by visualising auditory targets, recognition could have been based on non-structural information as well as structural information. In support of this interpretation, the same pattern of results was observed in Experiment 4, which employed unimodal trials. In turn, the relatively small effect of transformation in the AV condition of the present experiment can be attributed to possible ‘audiation’ strategies, in which audiated visual targets facilitated performance. As the facilitation effect was not strong enough to cancel out the performance advantage for inverse transformations, this suggests that participants were unable to audiate visual stimuli as effectively as they were able to visualise auditory stimuli (NB the same conclusion was made when comparing the hypothetical effects of visualisation and audiation strategies in Experiment 3). However, as has been stated before, interpretations based on these visualisation and audiation strategies are highly speculative. Nevertheless, they are useful to discuss because they demonstrate how the results could have been obscured by psychological processes that were not properly accounted for, despite the experimenter's efforts to isolate the specific psychological processes through careful control of all aspects of the experimental design.

Looking more generally at the effects of modality, performance was largely better in the VA condition, though on closer analysis the modality effect was only observed when participants were responding to retrograde transformations. When interpreted within the SPS framework, the absence of a modality effect for inverse transformations may be explained by the fact that recognition in both modalities involved an inversion of ordinal relations on equivalent supramodal scalar dimensions, suggesting that structural information was abstracted equally efficiently from auditory and visual stimuli. In contradiction, the VA advantage when responding to retrograde transformations would suggest that structural information was abstracted more efficiently from visual stimuli (assuming recognition was based on mental transformation of structural information abstracted from the standard). However, when taking into account possible 'visualisation' and 'audiation' strategies when recognising targets under retrograde transformation, the VA advantage can be attributed to participants being better able to visualise auditory targets than they were able to audiate visual targets. It should be noted that this pattern of results was very similar to those observed in Experiment 3, which also involved cross-modal trials. Taking into account the fact that visual stimuli were presented differently in both experiments, and the potential issues surrounding visualisation and audiation strategies that have been raised, it is difficult to draw any definitive conclusions from this consistency.

Finally, the exploratory analysis of the effects of music training on d' scores revealed that participants with some music training were generally better at detecting transformations than participants with no music training. This was interesting because music training had not been linked to better performance in Experiment 3. Taken in isolation, it suggests that music training confers advantages in the processing of structural transformation tasks, regardless of the sensory modality from which structural information has been abstracted. As ever, any findings from the exploratory analysis on music training must be viewed with caution, as the different groups were not carefully controlled.

In conclusion, the findings from Experiment 5 partially supported the 1½-D hypothesis. Even though auditory and visual stimuli were presented in cross-modal trials, and therefore recognition should have been based exclusively on the processing of structural information represented in 1½-D supramodal pattern spaces, it was possible that participants were able to use alternative strategies that obscured the results. To overcome this possibility, cross-modal experiments would need to present stimuli in such a way that makes it difficult for participants to learn to associate specific pitches with specific spatial positions. Experiment 6 attempted to do this by presenting stimuli in different pitch and spatial ranges (rather than presenting stimuli with the same limited range of pitch and spatial positions in all trials), and by transposing target patterns.

5.4 Experiment 6: Hybrid trials

The aim of Experiment 6 was to examine the 1½-D hypothesis in a hybrid experiment that combined unimodal and cross-modal conditions. In each trial, the standard was either an auditory (AS condition) or a visual pattern (VS condition), and the target was bimodal (i.e. auditory and visual patterns were presented simultaneously). In addition, target patterns were transposed so that they began on different pitches and at different spatial positions to standard patterns.

The results of the previous unimodal experiments reported in this thesis (Experiments 2 and 4) failed to support the 1½-D hypothesis, which predicted that inverse transformations would be recognised more successful than retrograde transformations. However, transformations were applied without transposition, which meant that, when stimuli corresponded to a 1½-D supramodal pattern space, recognition could be based on the processing of structural information and additional non-structural information. It remains to be seen whether support for the 1½-D hypothesis can be found when non-structural information is not available.

In previous cross-modal experiments (Experiments 3 and 5), when the 1½-D hypothesis has been tested it has received some support in AV conditions but not in VA conditions. In discussion of these results, it was argued that 'visualisation' and 'auditation' strategies may have interfered with the structural processes under investigation. For example, in Experiment 5 participants may have learnt to associate the pitches of tones with the spatial positions of visual objects, allowing them to make use of non-structural cues in a similar fashion to that described by the transposition hypothesis. Therefore, a better test of the 1½-D hypothesis in cross-modal conditions would present stimuli in a way that makes it difficult to use these alternative strategies.

The present experiment sought to address the above issues by presenting auditory or visual standard patterns followed by bimodal target patterns that were transposed to begin on different pitches or at different spatial positions. As usual, target patterns were either related to the standard under transformation (inverse or retrograde) or were unrelated. The modality of standards was manipulated (as opposed to targets) because previous experiments reported in this thesis have consistently supported that assumption that participants mentally transform the standard in anticipation of the target. Therefore, this design theoretically required the mental transformation of structural information abstracted from auditory or visual stimuli, which had to be compared with a bimodal stimulus. According to the SPS framework, both auditory and visual stimuli would be represented in a 1½-D supramodal pattern space, and the 1½-D hypothesis predicted that inverse transformations would be processed more effectively than retrograde transformations.

When comparing mental transformations to bimodal targets it was possible that participants would choose to attend either to auditory or visual information (or to both simultaneously), depending on the context. For example, when targets follow an auditory standard they might attend to auditory information, and when targets follow a visual standard they might attend to visual information. Alternatively, they might only attend to visual information, irrespective of the modality of the standard. Either way, it did not matter as both auditory and visual information corresponded to the same structural representation. The most important point was that, due to the transposition of bimodal targets, recognition could not be based on the processing of non-structural information. Furthermore, transposition also made it difficult for participants to use any ‘visualisation’ or ‘audiation’ strategies.

In summary, Experiment 6 tested the 1½-D hypothesis, which predicted that recognition performance would be best for inverse transformations in both modality conditions. The transposition of target stimuli ensured that recognition could not be based on non-structural information, which may have facilitated the recognition of retrograde transformations in previous experiments. In addition, transposition made it difficult for participants to use alternative ‘visualisation’ and ‘audiation’ strategies that may have obscured the results of previous experiments.

5.4.1 Methods

5.4.1.1 *Participants*

36 students from the University of Roehampton took part in Experiment 6 (female = 27, male = 9; mean age = 29.56 years, SD = 6.19). All had normal hearing and normal or corrected-to-normal vision. Three participants were left-handed, and the remainder were right-handed. 27 of the participants reported having received some music training (mean = 4.86 years). They all received course credit for their participation.

5.4.1.2 ***Stimuli***

The pattern structures used in previous recognition experiments were redesigned in Experiment 6. This was done for two reasons. First, it was noted that unrelated patterns used in the previous experiments could share the same contour structure. For example, the patterns ACABC and ABABC are *transformationally distinct*, and were classed as being unrelated in the previous experiments. However, whilst they have different structures at the interval level (+2,−2,+1,+1 and +1,−1,+1,+1), they share the same structure at the ordinal level (UDUU and UDUU). As Dowling (1972) has shown, melodic transformations that have different interval structure but share contour structure are frequently mistaken for identical transformations. Therefore, though the probability was very low, it is possible that on occasion standard patterns were paired with randomly selected unrelated targets that shared the same contour as a corresponding related target pattern. This would clearly have been problematic, as the participant might have mistakenly indicated that it was a related target. Second, it was anticipated that transposing target patterns would make an already difficult task even more difficult, so newly designed patterns were designed to be less complex relative to those used in previous experiments. To this end, new pattern structures consisted of change in ordinal relations only, with no variation in interval information. Also, patterns used in the experiment included ternary (as used in the previous experiments) and binary structures.

5.4.1.2.1 Pattern structure

The generation of patterns in Experiment 6 took as its starting point all possible combinations of the scalar-temporal relations ‘downwards’ (D), ‘same’

(S), and 'upwards' (U) in a length 4 string (e.g. UDSU, SUUD). The generation of pattern structures is described below. 81 patterns were generated using the above parameters. 62 of these were discarded as they did not meet the following criteria. First, all patterns were required to be *transformationally distinct* from one another. This meant that they could not be equivalent to an inverse, retrograde or retrograde inverse transformation of another pattern in the set. Second, all patterns were required to produce at least one pattern variation under transformation (this criterion was only relevant for pattern SSSS). Of the remaining set of 24 transformationally distinct patterns, 8 produced only one variation under all types of transformation, so were discarded. The result was a pool of 16 patterns that produced variations under all types of transformation (inverse, retrograde and retrograde inverse), providing four *transformational variants* of the same basic pattern. 10 patterns were randomly selected for use as standard and related target stimuli with the following constraints: 1) only binary (two scalar values) or ternary (three scalar values) patterns were selected, 2) an equal proportion of binary and ternary patterns were represented. For each of the 10 selected patterns, one of their transformational variants was randomly selected as the standard. The transformational variants produced by retrograde inverse transformation were discarded. See Table 5.1 for all standard and related target patterns.

Table 5.1

Patterns used in Experiment 6

Standard	Inverse		Retrograde	
	Related	Unrelated	Related	Unrelated
UDSS	DUSS	DUDS	SSUD	SDUD
DUDS	UDUS	UDSS	SUDU	SSDU
SSDS	SSUS	SDUS	SUSS	SUDS
DSSS	USSS	USDS	SSSU	SDSU
SUSD	SDSU	SSSU	USDS	USSS
USUD	DSDU	DSUU	UDSD	UUSD
SDDU	SUUD	SUDD	DUUS	DDUS
DSUU	USDD	USUD	DDSU	DUSU
SSUU	SSDD	SUDD	DDSS	DDUS
UUDU	DDUD	DUUD	DUDD	DUUD

Note. D = downwards; S = same; U = upwards

Unrelated patterns were created by altering one of the two middle scalar-temporal relations of the corresponding related target (see Table 5.1). This was carefully controlled to try and avoid using unrelated targets that were related to other standard and related target stimuli already being used in the experiment. They were also controlled to ensure that the unrelated target pattern comprised the same number of different scalar values as the corresponding standard/related target. This approach was taken to control the similarity between standard and unrelated targets in the experiment. Due to the limited size of the pattern set, occasionally it could not be avoided that an unrelated pattern was the same as or a transformational variant of another pattern used in the experiment. This was deemed acceptable, as the nature of the relationships between patterns presented within a single trial was more important than the nature of the relationships between patterns presented in different trials.

5.4.1.2.2 Auditory stimuli

Auditory stimuli were 5-tone pitch sequences. They were monophonic, isochronous, and composed from a 5-note equal temperament scale. All standard patterns began on the same pitch (f_0 520.00 Hz) and a scalar-temporal relation of U or D equated to an interval of one scale step. Target stimuli began on a pitch that was transposed 1 and a half scale steps either above or below the beginning pitch of the standard (f_0 422.37 Hz or f_0 640.20 Hz). This particular transposition distance was chosen to be roughly consistent with the transposition distances adopted in previous melody recognition experiments (Bartlett & Dowling, 1980; Cuddy & Cohen, 1976; Dowling & Fujitani, 1971; Dyson & Watkins, 1984; Lee, Janata, Frost, Martinez, & Granger, 2014; Miyazaki, 2004), and ensured that retrograde targets could not share the same pitch content as the standard (as was the case in Experiments 2 and 4).

Duration of tones was 350ms (with 10ms linear onset and offset ramps) and there was a 150ms ISI between the offset of an antecedent tone and the onset of a consequent tone. The IOI between each successive tone was 500ms. All tones had triangle waveforms, were generated using NCH Tone Generator version 3.02 (NCH Software), and edited using WavePad Sound Editor Masters Edition version 5.02 (NCH Software). WavePad was also used to arrange tones into sequences, which were digitally recorded as .wav file type (sample size 16 bit, sample rate 44 kHz, format PCM uncompressed, mono).

5.4.1.2.3 Visual stimuli

As in Experiments 4 and 5, each visual stimulus was a sequence of black bar segments presented at different vertical positions. The first segment of each standard visual stimulus was aligned to the centre of the display screen, and subsequent segments were presented at different positions along the central axis. One interval step on the scale dimension corresponded to a vertical visual angle of 1.35° (measured from the centre of segments). The first segment of target stimuli was presented the equivalent of one and a half scale steps above or below the centre of the screen, which was equal to a visual angle of 2.03°. As with Experiments 4 and 5, all segments of visual stimuli were presented on a vertical axis at the centre of the display screen.

The timing of events in the visual stimuli was identical to the timing of events in the auditory stimuli (though no equivalent of the 10ms onset/offset ramp was applied to the presentation of segments). Black bar segments were drawn using Adobe Illustrator and animated using Final Cut Pro.

5.4.1.2.4 Stimuli presentation in E-Prime

Auditory, visual and bimodal stimuli were produced into movie files for presentation in E-Prime. Audio-only, visual only, and audio-visual movie files were rendered using Final Cut Pro and exported as QuickTime format movie files. These then had to be converted to WMV format using MPEG Streamclip (version 1.9.2) to be compatible with the E-Prime software that was used to implement the experiments.

5.4.1.3 ***Design and procedure***

The experimental design and procedure were identical to Experiment 2, 3, 4 and 5 with the following exceptions. In half of the trials the standard pattern was auditory and in the other half visual. In all trials the target was bimodal (auditory and visual stimuli were presented simultaneously). The within-subjects factor of modality had two levels (auditory standard [AS], visual standard [VS]). Altogether there were 6 experimental conditions – two levels of modality (AS, VS), two levels of relatedness (related, unrelated), and two levels of transformation (retrograde, inverse) embedded into the former relatedness level. Consequently, the proportion of trials per condition was as follows: 1) AS, related, retrograde (ASRR) = 12.5%; 2) AS, related, inverse (ASRI) = 12.5%; 3) AS, unrelated (ASU) = 25%; 4) VS, related, retrograde (VSRR) = 12.5%; 5) VS, related, inverse (VSRI) = 12.5%; 6) VS, unrelated (VSU) = 25%.

10 standard patterns were used in the experiment. Each pattern was presented once in related conditions and twice in unrelated conditions, making a total of 80 trials per experimental session. The experimental session was divided into 40-trial blocks containing equal proportions of AS and VS trials and equal proportions of related and unrelated trials. Participants only had to recognise one type of transformation per block, therefore one block contained all the related inverse trials, and the other block contained all related retrograde trials. Each block was further sub-divided by modality into 20-trial sub-blocks, containing 10 related trials and 10 unrelated trials each. The order of blocks was counterbalanced between participants. The presentation order of sub-blocks within each block was randomised, as was the order of trials within each sub-block.

5.4.2 Results

All participants performed above chance level (overall error rates were less than 50%). Paired-samples *t*-tests were run on block order (first block, second block) to examine any effects of learning on overall percent error (PE) and response time (RT). There was no significant effect of block order on PE, $t(35) = -0.88$, $p = .194$, one-tailed. However, there was a significant effect of block order on RT, $t(35) = 3.09$, $p = .002$, one-tailed, with slower RT in the first block (M = 1039.15, SD = 380.87) than in the second block (M = 930.57, SD = 406.32).

5.4.2.1 ***Error data***

5.4.2.1.1 All trials

Overall PE was 27.92. This is in keeping with the previous recognition experiments reported in this thesis in which participants made on average 26% error. Mean results for responses to inverse, retrograde and unrelated targets in both modality conditions are displayed in Figure 5.8.

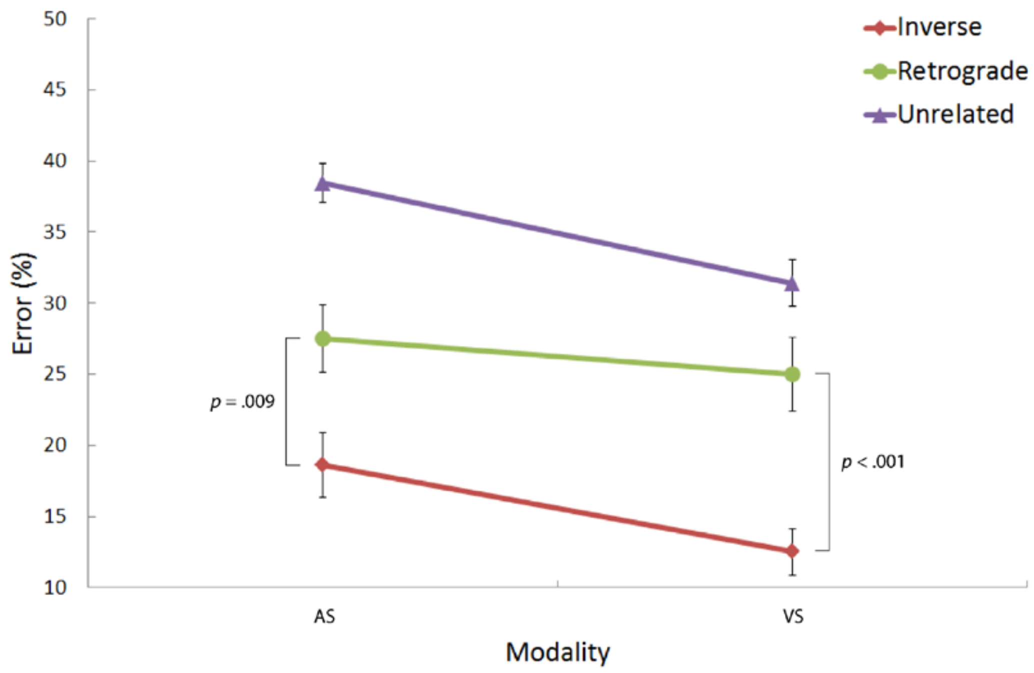


Figure 5.8. Experiment 6: Mean PE in target conditions, plotted as a function of modality. Significance values for simple effects of transformation were obtained from the analysis on arcsine-transformed data (NB simple effects of modality are not displayed). Error bars indicate the standard error of the mean (SEM ± 1).

A 2 x 2 ANOVA was performed on arcsine-transformed PE data with relatedness (related, unrelated) and modality (AS, VS) as the within-subjects factors. The analysis revealed a highly significant main effect of relatedness, $F(1,35) = 46.12$, $MSE = .03$, $p < .001$, $\eta_p^2 = .57$. Participants made more errors when responding to unrelated targets ($M = 34.93$, $SD = 13.97$) than related targets ($M = 20.90$, $SD = 11.45$). The main effect of modality was significant, $F(1,35) = 6.93$, $MSE = .02$, $p = .013$, $\eta_p^2 = .17$, with more errors being made in the AS ($M = 30.76$, $SD = 12.52$) than in the VS condition ($M = 25.07$, $SD = 11.04$). The interaction between relatedness and modality failed to reach significance, $F(1,35) = 0.27$, $MSE = .02$, $p = .609$, $\eta_p^2 = .008$.

5.4.2.1.2 Related trials only

The results of the above analysis permitted the performance of a further 2 x 2 ANOVA that examined the within-subjects effects of modality (AS, VS) and transformation (inverse, retrograde) for related targets only. In the absence of data from responses to unrelated targets, the main effect of modality failed to reach significance, F(1,35) = 2.08, MSE = .05, p = .158, η_p^2 = .06. Visual inspection of Figure 5.8 shows that PE was lower for inverse transformations in both modality conditions, and there was a highly significant main effect of transformation, $F(1,35) = 18.14$, $MSE = .05$, $p < .001$, $\eta_p^2 = .34$. This finding supported the 1½-D hypothesis. The interaction between modality and transformation failed to reach significance, $F(1,35) = 1.10$, $MSE = .02$, $p = .302$, $\eta_p^2 = .03$.

Pairwise comparisons were carried out to examine the simple effects of transformation in both modality conditions. They confirmed that the effect of transformation was significant in both the AS (*MD* = 8.89, *SE* = 3.42; *p* = .009) and the VS conditions (*MD* = 12.50, *SE* = 2.86; *p* < .001). Further pairwise comparisons were run to examine the simple effects of modality in both transformation conditions. In the inverse condition the mean difference was approaching significance (*MD* = 6.11, *SE* = 2.12; *p* = .055), with less error being made in the VS condition. The mean difference failed to reach significance in the retrograde condition (*MD* = 2.50, *SE* = 3.34; *p* = .577).

5.4.2.2 ***RT data***

5.4.2.2.1 All trials

The same analysis was repeated in full on log-transformed RT data. Overall mean RT was 979.26ms, which was similar to the response times in the previous recognition experiments (mean 947.84ms). Mean results for responses to inverse, retrograde, and unrelated targets in both modality conditions are displayed in Figure 5.9.

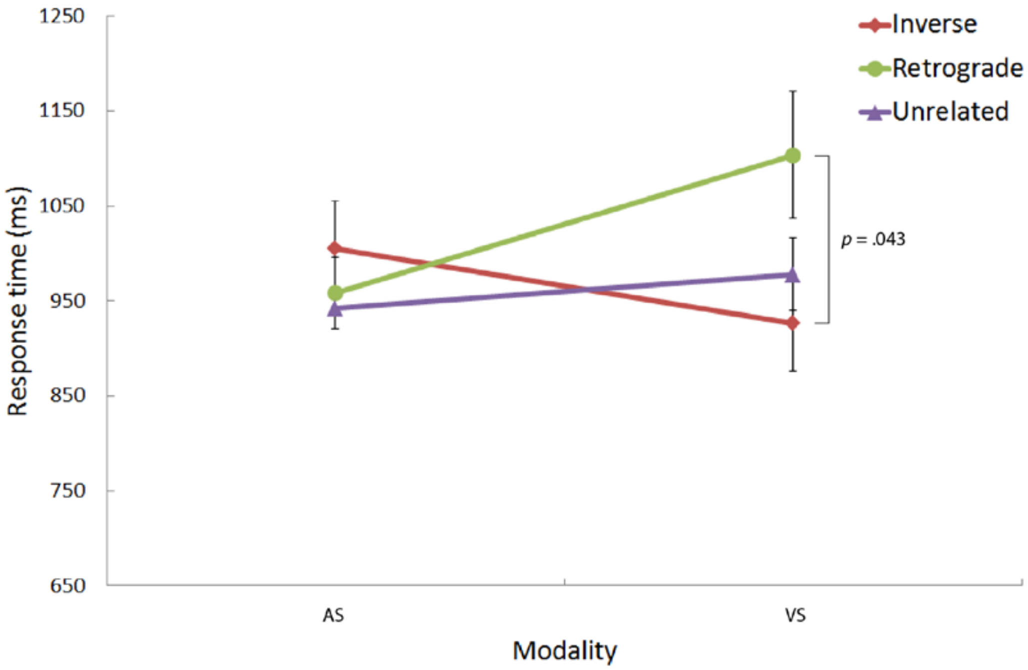


Figure 5.9. Experiment 6: Mean RT in target conditions, plotted as a function of modality. Significance values for simple effects of transformation were obtained from the analysis on log-transformed data. Error bars indicate the standard error of the mean (SEM ± 1).

A 2 x 2 ANOVA was performed to examine the effects of relatedness and modality. The analysis failed to produce any significant results. An effect of relatedness had been observed on RT data (and indeed on PE data) in all previously reported recognition experiments – it is not clear why the effect failed to reach significance here.

5.4.2.2.2 Related trials only

Further analysis performed on related targets only. This was justified because, despite finding no significant effects when the above analysis was performed, responses to unrelated targets were found to be significantly slower in all previous experiments. In addition, though the present experiment's RT data failed to reveal a significant effect of relatedness, analysis of the error data above did show a significant effect. The 2 x 2 ANOVA revealed that the main effect of modality failed to reach significance, $F(1,35) = 0.01$, $MSE = .09$, $p = .909$, $\eta_p^2 < .001$. Visual inspection of Figure 5.9 shows that in the AS condition mean RT was faster for retrograde transformations, contrary to the prediction made by the 1½-D hypothesis. In the VS condition, on the other hand, mean RT was faster for inverse transformations. The main effect of transformation did not reach significance, $F(1,35) = 1.16$, $MSE = .21$, $p = .288$, $\eta_p^2 = .03$, failing to support the 1½-D hypothesis. There was a significant interaction between transformation and modality, $F(1,35) = 5.56$, $MSE = .07$, $p = .024$, $\eta_p^2 = .14$.

Pairwise comparisons were run to examine the simple effects of transformation in both modality conditions. In the AS condition the mean difference failed to reach significance (MD = 46.70, SE = 74.90; p = .814). However, in the VS visual condition the mean difference was significant (MD = 177.42, SE = 95.16; p = .043). This result provided partial support for the 1½-D hypothesis. Further comparisons were carried out to examine the simple effects of modality but the mean difference failed to reach significance in both retrograde (MD = 145.67, SE = 64.90; p = .141) and inverse conditions (MD = 78.45, SE = 64.00; p = .115).

5.4.2.3 ***Signal detection analysis***

Further analysis was carried out using signal detection theory.

5.4.2.3.1 Sensitivity to the signal

Mean d' was 1.21. This was low compared with the average of 1.36 observed across previous experiments, though not as low as the mean d' observed in Experiment 2 (1.06). A 2 x 2 within-subjects ANOVA was carried out to examine the effects of modality (AS, VS) and transformation (inverse, retrograde) on the detectability of related targets. There was a significant main effect of modality, $F(1,35) = 8.76$, $MSE = .38$, $p = .006$, $\eta_p^2 = .20$, with detectability being better in the VS ($M = 1.37$, $SE = .12$) than in the AS condition ($M = 1.06$, $SE = .13$). Visual inspection of Figure 5.10 shows that detection was better for inverse transformations in both modality conditions, and there was a significant main effect of transformation, $F(1,35) = 9.01$, $MSE = .47$, $p = .005$, $\eta_p^2 = .21$. This result is in agreement with PE data and provided further support for the 1½-D hypothesis. However, the interaction between transformation and modality was approaching significance, $F(1,35) = 3.81$, $MSE = .37$, $p = .059$, $\eta_p^2 = .10$.

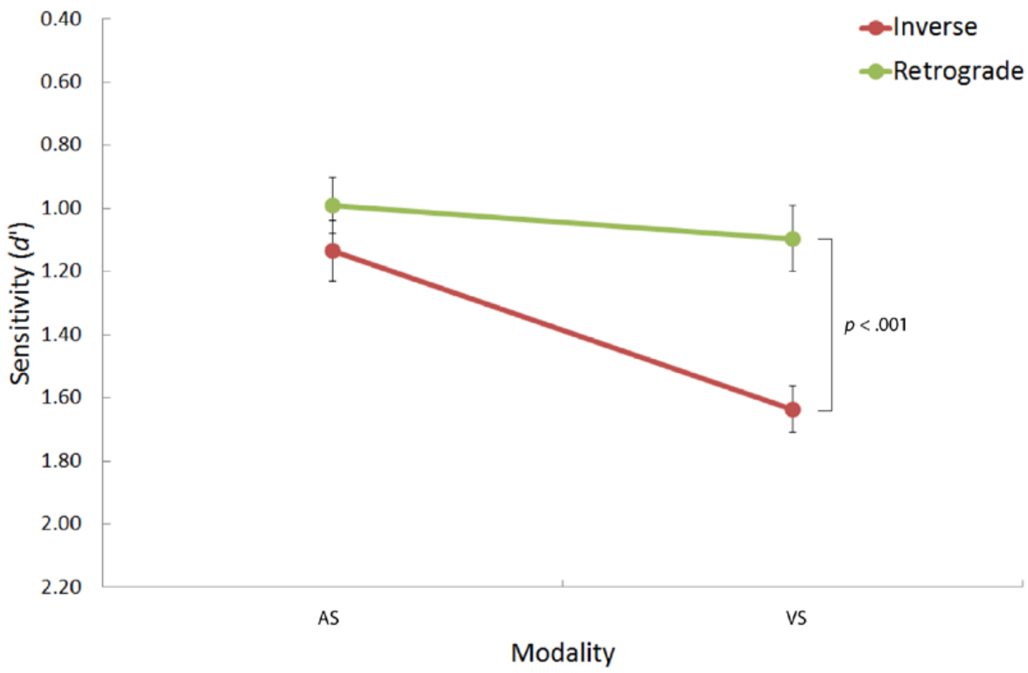


Figure 5.10. Experiment 6: Mean *d'* in transformation conditions, plotted as a function of modality (NB simple effects of modality are not displayed). The scale along the y-axis has been inverted so that results can be more easily compared with figures displaying PE data. Error bars indicate the standard error of the mean (SEM ± 1).

The interaction reflected a significant sensitivity advantage for inverse targets in the VS condition. Pairwise comparisons examined the simple effects of transformation in both modality conditions. They revealed that while the mean difference was highly significant in the VS condition (*MD* = 0.54, *SE* = .15; *p* = .001), it failed to reach significance in the AS condition (*MD* = 0.15, *SE* = .15; *p* = .353). Therefore, these results provided only partial support for the 1½-D hypothesis. Further pairwise comparisons were run to examine the simple effects of modality. They revealed a highly significant mean difference in the inverse condition (*MD* = 0.50, *SE* = .13; *p* < .001), with detection being better in the VS condition. In the retrograde condition the mean difference failed to reach significance (*MD* = 0.12, *SE* = .16; *p* = .506).

5.4.2.3.2 Response bias

Mean c was -0.20. The same score was observed in Experiment 5, and it was more liberal than the average score observed in previous experiments (-0.14). A two-way ANOVA was carried out to investigate participants' tendency to respond with either 'related' or 'unrelated'. There was a highly significant effect of transformation, $F(1,35) = 12.51$, $MSE = .09$, $p = .001$, $\eta_p^2 = .26$, with responses being increasingly liberal in the inverse condition ($M = -0.29$, $SD = .26$) compared with retrograde ($M = -0.11$, $SD = .22$). There was no significant main effect of modality, and there was no significant interaction between the variables.

5.4.2.4 ***Music training analysis***

27 participants were allocated to the 'some training' group, and 9 participants were allocated to the 'no training' group. A Pearson's correlation analysis revealed two significant correlations between the amount of music training reported by participants in the some training group and performance when recognising inverse transformations in the AS condition. Increasing levels of music training was associated with decreasing error rates, $r = -.41$, $p = .030$, and increasing sensitivity, $r = .46$, $p = .015$.

A series of 2 x 2 x 2 mixed ANOVAs, with the within-subjects factors modality (AS, VS) and transformation (inverse, retrograde) and the between-subjects factor music training (some training, no training), were run on PE, RT, d' and c data. The analysis on PE (arcsine transformed) data revealed a significant main effect of music training, $F(1,34) = 5.82$, $MSE = .12$, $p = .021$, $\eta_p^2 = .15$. Participants in the some training group ($M = 18.61$, $SD = 10.41$) made less error than participants in the no training group ($M = 27.78$, $SD = 12.28$). There was also a significant interaction between transformation and music group, $F(1,34) = 8.55$, $MSE = .04$, $p = .006$, $\eta_p^2 = .20$. Pairwise comparisons were run to examine the simple effects of music training in both transformation conditions and revealed a highly significant mean difference between music training groups in the inverse condition ($MD = 17.78$, $SE = 4.18$; $p = .001$). Participants with some training ($M = 11.1$, $SD = 9.23$) made less error than participants with no training ($M = 28.89$, $SD = 14.95$). The analysis on RT (log-transformed) failed to reveal any significant results.

The analysis on *d′* data revealed a significant interaction between transformation and music group, $F(1,34) = 4.85$, $MSE = .42$, $p = .035$, $\eta_p^2 = .13$. Pairwise comparisons were run to examine the simple effects of music training in both transformation conditions. A significant mean difference between music training groups was revealed in the inverse condition ($MD = 0.80$, $SE = 0.30$; $p = .010$), with participants in the some training group ($M = 1.59$, $SD = 0.78$) being better at detecting inverse transformations than participants in the no training group ($M = 0.79$, $SD = 0.73$). The analysis on *c* data also revealed a significant interaction between transformation and music group, $F(1,34) = 7.21$, $MSE = .08$, $p = .011$, $\eta_p^2 = .18$. Pairwise comparisons were run to examine the simple effects of music training in both modality conditions. They revealed a significant mean difference between music groups in the inverse condition ($MD = 0.21$, $SE = 0.10$; $p = .033$). When detecting inverse transformations, both music training groups were liberal with their responses, but participants with some training ($M = -0.34$, $SD = 0.27$) were more liberal than participants with no training ($M = -0.13$, $SD = 0.14$).

5.4.3 Discussion

The key finding in Experiment 6 was that for the first time in all recognition experiments reported in the present and previous chapter, the 1½-D hypothesis received some support in both modality conditions. In general, recognition performance was better for inverse transformations. What makes this finding particularly convincing is the fact that the experiment was controlled to eliminate the possibility that recognition could be based on anything but the mental transformation of structural information. This finding is in agreement with the results reported by Dowling (1972), who investigated the recognition of inverse and retrograde transformations of transposed melodies. The present experiment extends these findings by demonstrating that the same pattern of results is obtained when participants have to recognise transformations of structurally analogous visuo-spatial patterns.

More than this, the success of the 1½-D hypothesis supported the assumptions of the SPS framework, and shows that under the appropriate conditions, auditory and visual patterns inhabit the same structural "space". According to the framework, the auditory and visual stimuli used in the present experiment can be represented in a 1½-D supramodal pattern space, constructed from a scalar and a temporal dimension. Inverse transformations were recognised more successfully than retrograde transformations because they require an inversion of ordinal relations on a scalar dimension. In contrast, retrograde transformations require an inversion of ordinal relations on a temporal dimension, which is harder to process due to the dimension's inherent directionality.

Despite the general support, the hypothesis was not fully supported by the analysis on RT and d' data. In both cases, the 1½-D hypothesis was supported in the VS condition, but not in the AS condition. In the AS condition, mean RT was actually slower for inverse transformations, though the difference between transformation conditions was not significant. In agreement with the 1½-D hypothesis, mean d' was higher for inverse transformations, indicating better detection. But the difference between transformation conditions also failed to reach significance. This raises the question, why did the analysis on PE data reveal an advantage for inverse transformations in the AS condition, but the analysis on d' data did not? The answer can be found by looking at the proportions of hits and false alarms. Although the proportion of hits was higher for inverse transformations (inverse = 0.80, retrograde = 0.73), the proportion of false alarms was also higher (inverse = 0.41, retrograde = 0.36). Therefore, it is possible that the higher proportion of hits reflected participants' tendency towards a more liberal response bias when identifying inverse transformations, rather than better perception. With this in mind, the large performance advantage for inverse transformations in the AS condition, revealed by the analysis on PE data, should be viewed with some caution.

A possible explanation of the inconsistent performance advantage for inverse transformations in the AS condition becomes apparent when also taking into account the effects of modality on performance. The only effect of modality was revealed by the analysis on d' data – the detection of inverse transformations was better in the VS than the AS condition. A visual advantage has been observed in previous experiments, and two possible explanations have been proposed: 1) visualisation strategies were employed when recognising retrograde transformations; 2) structural information was abstracted more efficiently from visual stimuli. In the present experiment, visualisation strategies were unlikely to be deployed due to the transposition of targets. Also, visualisation strategies have been implicated in the recognition of retrograde transformations – the present modality effect was observed for inverse transformations; therefore, an explanation involving visualisation strategies can be ruled out.

This leaves the possibility that structural information was abstracted more efficiently from visual stimuli. If this were the case, why was the modality effect not observed for both types of transformation? The answer may lie in the fact that different transformations require an inversion of ordinal relations on different types of supramodal dimension – inverse on the scalar and retrograde on the temporal. Representations on the scalar dimension correspond to the relative pitch of tones or the relative vertical height of objects, whilst representations in the temporal dimension correspond to the relative timing of auditory or visual events. Therefore, the visual advantage may reflect more efficient abstraction of structural information from the visual spatial dimension compared to the auditory pitch dimension. In turn, the absence of a modality effect on the recognition of retrograde transformations would imply that temporal structure was abstracted from auditory and visual stimuli with equal efficiency.

Whilst this explanation is inviting, it contradicts the traditional view that the auditory system is superior in temporal processing (Collier & Logan, 2000; Gault & Goodfellow, 1938). However, the stimuli used in the present experiment were isochronous and therefore temporally very simple. Perhaps with more complex rhythmic stimuli an auditory advantage would be observed for retrograde transformations, resulting from more efficient abstraction of temporal structure from auditory stimuli. Future experiments could investigate this by manipulating scalar and temporal complexity independently – increasing scalar complexity would predict an increasing modality effect on inverse transformations, with better performance when structural information has been abstracted from visual stimuli; increasing temporal complexity would predict an increasing modality effect on retrograde transformations, with better performance when structural information has been abstracted from auditory stimuli.

Finally, the exploratory analysis on the effects of music training revealed some interesting results. Participants with some music training performed better than participants with no music training when recognising inverse transformations, but not retrograde transformations (though there was no effect of music training on RT). This did not interact with the modality condition. This would suggest that music training conferred a processing advantage that was not limited to structural information abstracted from auditory stimuli. This on its own supports the notion of supramodal structural representations and processes, as experience in one sensory domain benefitted processing in the other sensory domain. It is not clear why music training should have an effect on the processing of inverse transformations but not retrograde transformations. However, when thinking about how both transformations correspond to different supramodal dimensions, this suggests that music training improves processing of structural information on scalar dimensions but not temporal dimensions. Perhaps an effect of music training would be revealed for retrograde transformations when more complex rhythmic stimuli are used.

In conclusion, the results of Experiment 6 provide some support for the 1½-D hypothesis in different modality conditions, and therefore validate the assumptions of the SPS framework. However, the 1½-D hypothesis received less support in the AS condition, and further research is required to confirm the findings. When interpreted within the SPS framework, it is possible that the partial support for the hypothesis in the AS condition reflects superior abstraction of structural information from visual spatial dimensions, in comparison to the dimension of auditory pitch.

5.5 General discussion

The general aim of Experiments 4, 5 and 6 has been to examine the possibility that structural information, abstracted from auditory and visual stimuli, is represented in a shared 1½-D supramodal pattern space. An additional aim was to explore the possibility that a supramodal mechanism (or mechanisms) is responsible for the processing of inverse and retrograde transformations of sequential pattern structure. Following the SPS framework, outlined in Chapter 2, the 1½-D hypothesis was tested in all experiments. The hypothesis was not supported in some experimental conditions, when the recognition of retrograde transformations could also be based on non-structural information. When the availability of non-structural information was removed, the 1½-D hypothesis was mostly supported (see Table 5.2 for a summary).

Table 5.2

Chapter 5: Summary of experiments, hypotheses and results

Experiment	Modality	Hypothesis tested	Result
4	A	1½-D	Not supported
	V	1½-D	Not supported
5	AV	1½-D	Some support
	VA	1½-D	Not supported
6	AS	1½-D	Some support
	VS	1½-D	Supported

Note. A = auditory; V = visual; AV = auditory-visual; VA = visual-auditory; AS = auditory standard; VS = visual standard

The main 1½-D hypothesis predicted that inverse transformations would be recognised more successfully than retrograde transformations, irrespective of the sensory modality in which patterns were presented. This hypothesis assumed that recognition would be based exclusively on the processing of structural information. The auditory and visual stimuli used in the experiments reported in the present chapter were both dimensionally equivalent and corresponded to a 1½-D supramodal pattern space, constructed from a scalar and a temporal dimension. The scalar dimension represents the relative pitch height of tones or the relative spatial position of visual objects on the vertical axis. The temporal dimension represents the relative timing of auditory or visual events. Inverse transformations of patterns represented in this 1½-D space require an inversion of ordinal relations on a scalar dimension, whilst retrograde transformations require an inversion of ordinal relations on a temporal dimension. Retrograde transformations are hypothesised to be harder to process because of the inherent directionality of the temporal dimension, on which structural relations must be inverted.

The results of Experiment 4 failed to support the 1½-D hypothesis. As discussed in Section 5.2.3, the results may have reflected a facilitation effect of non-structural information when recognising retrograde transformations. In an attempt to control for this possibility, Experiment 5 presented patterns cross-modally and Experiment 6 transposed target patterns. The results of Experiment 5 provided some support for the 1½-D hypothesis in the AV condition (when auditory standards were followed by visual targets) but not in the VA condition (when visual standards were followed by auditory targets). However, as discussed in Section 5.3.3, it was possible that participants were able to use alternative 'visualisation' and 'audiation' strategies that facilitated the recognition of retrograde transformations, obscuring the results. Although this explanation is speculative, the results of Experiment 6 seem to support it. The design of Experiment 6 made it difficult for participants to use these alternative strategies, and as a result the 1½-D hypothesis was largely supported.

When taken together, the pattern of results across all three experiments demonstrates that, when recognition can be based on the processing of non-structural information, inverse and retrograde transformations are recognised equally well. However, as recognition becomes more reliant on structural processing, a general advantage for inverse transformations emerges. This general trend is in line with the prediction made by the 1½-D and more broadly, with the assumptions of the SPS framework. Thus, although the results of the experiments reported in the present chapter do not provide conclusive support for the SPS framework, they offer some encouragement.

Chapter 6: Priming perception with structural transformations

6.1 Introduction

The main aim of the experiments reported in the present chapter was to investigate the perception of inverse and retrograde transformations of sequential pattern structure by hypothesised supramodal mechanisms. As in previous experiments, the focus was on auditory pitch sequences and structurally analogous visuo-spatial sequences. Unlike in previous experiments, a paradigm was adopted that neither explicitly instructed participants to compare patterns, nor provided any training on the types of transformation that might relate them. An additional aim was to explore the time courses of the hypothesised supramodal mechanisms involved in the processing of structural transformations.

In the experiments reported in Chapters 4 and 5 of this thesis, the perception of inverse and retrograde transformations was examined in a series of recognition experiments with different sensory modality conditions. The experiments demonstrated that participants are able to compare structural information from auditory and visual stimuli by applying relevant mental transformations. The main finding from these experiments was that, when auditory and visual stimuli corresponded to equivalent 1½-D supramodal pattern spaces (see the SPS framework, outlined in Chapter 2), and recognition was based on the processing of structural information, inverse transformations were largely easier to process than retrograde transformations. According to the SPS framework, this is because the perception of inverse transformations required a mental process that inverts ordinal relations on a scalar dimension, whereas the perception of retrograde transformations required a mental process that inverts ordinal relations on a temporal dimension. Inversions on the temporal dimension are harder to process because they are incompatible with the dimension's inherent directionality.

The short-term recognition paradigm that was used in these experiments required that participants were explicitly instructed to compare patterns, and also required that the experimenter demonstrate to participants the types of transformations that they had to identify. Therefore, all participants received some training on how to identify pattern relationships, and perception was based on a conscious awareness of the way in which patterns might be related. In a non-experimental context, the perceptual system must typically process structural regularities when no prior instruction or training has been provided. Whilst it is difficult to investigate pattern perception in a non-experimental context, it is possible to use a paradigm that avoids giving explicit instruction and training to participants, and is therefore arguably more ecologically valid.

The experiments reported in Chapters 4 and 5 clearly demonstrate that inverse and retrograde structural transformations can be perceived when participants have been informed of the types of transformation that might relate patterns. Nevertheless, even with training the task was not easy, as demonstrated by the relatively high error rate (the average error rate across all experiments was 26.43%, when chance performance would be 50%). Just because structural transformations are perceived under these conditions, it does not mean that they will be perceived when no previous instruction and training is provided. However, there is some limited evidence from melodic processing research to show that they are (Balch, 1981; Dienes & Longuet-Higgins, 2004). In a study by Balch (1981), participants were presented with two melodies and simply rated the second of the two for 'good continuation'. 'Good continuation' relates to the Gestalt principle which states that objects tend to be perceptually grouped when one follows the established configuration of the other (e.g. Koffka, 1935). The second melody was either a transformation of the first or it was unrelated. Inverse and retrograde transformations received significantly higher good continuation ratings than unrelated melodies, and it was concluded that these higher ratings were based on the perception of structural relationships between the transformed melodies.

A more recent experiment by Dienes and Longuet-Higgins (2004) has demonstrated that tone rows containing inverse and retrograde transformations can be recognised implicitly, but with some difficulty. The term implicit is used here to refer to the learning of rules and regularities solely through exposure to stimuli that follow a particular structure, and the application of this knowledge with little or no conscious awareness (Kuhn & Dienes, 2005). In a learning phase musician participants rated melodies that contained inverse and retrograde transformations for pleasantness. In a subsequent test phase they were told that the stimuli they had previously heard obeyed some set of rules, and that half of the stimuli that they were about to hear would obey the same rules and half would not. Only participants with routine exposure to and interest in music that frequently incorporates inverse and retrograde transformations into the compositional process (e.g. serialist music) were found to classify stimuli at a level that was significantly better than chance. Analysis of confidence ratings that accompanied responses suggested that participants believed they were guessing, regardless of how accurate they were, suggesting that perception of the transforms was implicit rather than explicit.

One problem with the above paradigms is that, although participants were not given any training regarding the types of transformation that might be applied to patterns, they were explicitly instructed to compare them. Another way of investigating the perception of structural transformations without having to train participants, but also without having to instruct them to compare patterns, involves the priming of expectations. According to Summerfield and Egner (2009), ‘expectations are brain states that reflect prior information about what is possible or probable in the forthcoming sensory environment’ (p.403). Importantly, expectations can guide attention in an environment, facilitating the perception of expected objects and events. In research addressing visual attention, a large number of behavioural studies have shown that spatial-, object- and feature-based attention can be primed by cues, resulting in faster and more accurate detection of expected targets (for reviews see Carrasco, 2011; Scholl, 2001). Classic work by Posner and colleagues (Posner, Nissen, & Ogden, 1978; Posner, Snyder, & Davidson, 1980; Posner & Snyder, 1975a, 1975b) typically employed a visual spatial cuing paradigm which requires participants to identify targets in a visual field full of distracting stimuli. The location of the target can be cued by an arrow placed at a fixation point or by placing a cue at the location of the subsequent target. When the cue indicates the correct position of the target, performance is facilitated. But when it indicates the incorrect location, performance is impaired.

This paradigm has been adapted to show that temporal attending (i.e. attending to particular moments in time) can also be primed in the visual, auditory, and tactile domains (for reviews see Correa, Lupiáñez, Madrid, & Tudela, 2006; Nobre, 2001). Furthermore, experiments have demonstrated that the detection of tones can be facilitated by prior knowledge of the to-be-detected tone (Greenberg & Larkin, 1968; Tanner & Norman, 1954). Early experiments employed a tone detection paradigm in which participants were required to identify which of two noise bursts contained an additional tone signal. Detection of a particular tone dropped to chance levels when its frequency was changed without informing participants (Tanner & Norman, 1954). However, detection of different frequencies was facilitated by cues that fell within a limited range of the to-be-detected signal frequency (Greenberg & Larkin, 1968). Collectively, the attention research reported above has been interpreted as demonstrating that expectancies based on prior knowledge can direct attention in space, time and frequency, and that attention has a limited bandwidth, so targets that fall outside that limited range are poorly detected.

Expectancies can also be based on more complex patterned contexts, and there have been a few studies that have investigated how simple patterns can influence perceptual responses in the visual and auditory domains (Dowling, Lung, & Herrbold, 1987; Dowling, 1973; Howard, O'Toole, Parasuraman, & Bennett, 1984; Johnston & Jones, 2006). In the auditory domain, Howard et al. (1984) extended the tone detection paradigm described above to include more complex pattern cues. In a series of experiments, participants were presented with trials in which two 12-tone melodies were accompanied by a background of continuous noise. In one of the melodies, the 11^{th} note had been removed, and the participants' task was to indicate which of the two melodies was incomplete. It was found that, for melodies that followed a simple ascending of descending pitch trajectory, tones were detected more accurately when they conformed to the continuous contour pattern. This result suggests that perception of the global context pattern cued expectations for the pattern to continue in the same way, and that the result of this was a dynamic focus of attention to a particular frequency range. An alternative approach was adopted by Dowling et al. (1987) who demonstrated that pattern-based expectancies not only guide attention in frequency/pitch but also in time. In a series of experiments participants were presented with a cue melody followed by a comparison that consisted of a melody interspersed with distractor tones. The melody embedded in the comparison was either the same as the cue, or a change had been made to one of its tones. The change applied to tones could affect its pitch (higher or lower than expected) and/or its temporal position (earlier or later than expected). Participants were significantly better at judging the pitch of tones when they fell within the expected pitch range and when presented at the expected time. The authors concluded that prior knowledge of a melody is used to form ‘expectancy windows’ defined by pitch and time.

Another relevant area of research is the investigation of the phenomenon of representational momentum (Hubbard & Bharucha, 1988; Johnston & Jones, 2006). Representational momentum refers to the tendency of participants to judge the final position of an event undergoing implied motion incorrectly. One explanation for this effect is based on expectancies associated with the exposure to patterned stimuli (Johnston & Jones, 2006). Hubbard and Bharucha (1988) investigated the influence of lower and higher-order pattern relationships on representation momentum in the visual domain. Participants were presented with a target circle that moved either horizontally or vertically at a constant velocity (lower-order pattern) and bounced back and forth within an implied frame (higher-order pattern). The task was to indicate the on-screen location at which the circle disappeared from view. (It disappeared either just before collision with the enclosing frame, at the point of collision, or post collision.) At post collision points, participants made errors that were consistent with the lower-order pattern of its linear motion, i.e. they were more likely to indicate a final position that was in the direction of motion, as opposed to its actual location of disappearance. However, at pre-collision and collision points participants' errors were consistent with the anticipation of a future change in direction, suggesting that expectancies were generated based on the higher-order periodic regularity extracted from the pattern of motion.

In an auditory equivalent of this study, Johnston and Jones (2006) presented participants with simple pitch patterns that were followed by probe tones. Participants judged whether the probe tones were the same pitch as the final tone in the pattern, or were tuned slightly lower or higher in pitch. In agreement with the findings of Hubbard and Bharucha (1988), it was demonstrated that pitch judgement errors were consistent with the extraction of lower order and higher order pattern relationships. For simple linear patterns of ascending or descending pitch, participants' errors were consistent with extrapolation of the implied trajectory of the pattern in pitch space (i.e. for ascending patterns, they were more likely to judge higher probe tones as being the same, and for descending patterns they were more likely to judge lower probe tones as the same). For more complex periodic motions, errors at period boundaries were consistent with anticipations for a change in pitch direction.

For pattern-based expectations to be generated, a structural invariant must be abstracted from the environment and then applied to a given context (e.g. Jones, 1990). Two special kinds of structural invariance are found in patterns under inverse and retrograde transformation. General evidence for expectancies based on the abstraction of patterns related under inverse and retrograde transformation can be found in the research on serial pattern learning. Classic work in this area has shown that when participants are exposed to a pattern, they can correctly predict future elements based on these global relations (Restle & Brown, 1970; Restle, 1970). It follows that, should the perceptual system be capable of abstracting the invariance that relates patterns under these transformations, this information will be used to generate expectations that guide attention and facilitate the perception of correctly anticipated events.

The experiments reported in the present chapter aimed to investigate whether inverse and retrograde transformations of sequential pattern structure are perceived without explicit instruction and training, by examining the facilitation effects of pattern-based expectancies in an indirect perceptual task. In order to do this, a structural priming paradigm was employed in which participants were presented with a prime pattern followed by a target pattern. Prime and target patterns were either related under isomorphic transformation (inverse, retrograde) or unrelated. In Experiment 7, both prime and target patterns were auditory pitch patterns. In Experiment 8, the prime was an auditory pitch pattern, but the target was a visuo-spatial pattern. Participants were given no prior information regarding the way in which prime and target patterns may or may not be related and were simply instructed to make a comparison between the final two events of target patterns – for auditory patterns this was a pitch comparison task in which they indicated whether the last tone was higher or lower than the preceding tone; for visual patterns this was a spatial comparison task in which they indicated whether the final object was higher or lower than the preceding object. The idea was that, should the global structural relationship between prime and target patterns be perceived, then specific expectations for how the target unfolds should be generated, guiding attentional resources in pitch space or in vertical space. In turn pattern events that conform to these expectations and therefore fall within attended-to 'expectancy windows' (Dowling et al., 1987) should be perceived more effectively, resulting in faster and more accurate responses.

The main hypothesis, called the transformation priming hypothesis, was that responses to events embedded in related targets would be facilitated when compared to responses to events embedded in unrelated targets, irrespective of the modality of prime and target patterns. This hypothesis was based on a number of factors. Firstly, as reported above there is some limited evidence that melodic transformations are perceived when no prior training has been given to participants (Balch, 1981; Dienes & Longuet-Higgins, 2004). Secondly, serial pattern learning research has demonstrated that global structural regularities, such as those described by inverse and retrograde transformation, can be used to correctly anticipate events in auditory and visual sequences (Fountain & Rowan, 1995; Jones & Zamostny, 1975; Kundey & Rowan, 2014; Restle, 1970, 1976). Thirdly, the SPS framework assumes that structural information, abstracted from auditory and visual stimuli, is represented in a supramodal pattern space and that any mechanisms responsible for the perception of structural transformations operate on these supramodal representations.

In addition to the main hypothesis, a more specific hypothesis was tested which concerned the relative perceptual salience of different transformations. According to the SPS framework, the stimuli used in the present experiment corresponded to representations in a 1½-D supramodal pattern space. Inverse transformations of such patterns require an inversion of ordinal relations on a scalar dimension, and retrograde transformations require an inversion of ordinal relations on a temporal dimension. One of the assumptions of the SPS framework is that inversions on the scalar dimension are easier to process than inversions on the temporal dimension, due to the latter dimension's inherent directionality (an assumption supported by the findings of Experiment 6, reported in Chapter 5). Thus, the 1½-D hypothesis predicted that the facilitation effect of structural relatedness should be greatest for targets under inverse transformation.

Finally, the time courses of processes involved in the perception of transformed pattern structure were also investigated by varying the inter-stimulus interval (ISI) between prime and target patterns. The manipulation of ISI in priming experiments has been used extensively in the study of language processing (e.g. Carter, Hough, Stuart, & Rastatter, 2011). The motivation for this line of investigation was based on the assumption that distinct cognitive processes follow distinct cortical pathways in the brain, and that these anatomical differences may be revealed by distinct patterns of behavioural data across the different time spans separating prime and target stimuli. There is a substantial amount of research that supports the general view that there are different functional pathways in the brain (discussed in Chapter 1, Section 1.7; Arnott, Binns, Grady, & Alain, 2004; D. J. K. Barrett & Hall, 2006; H. C. Barrett & Kurzban, 2006; Bullier, 2001; Hubel & Wiesel, 1968; Kaas & Hackett, 2000; Moerel, De Martino, & Formisano, 2014; Ungerleider & Mishkin, 1982; Van Essen & Maunsell, 1983). However, no cognitive or neurocognitive theoretical models currently exist that specifically address the possible time courses of processes involved in the perception of inverse and retrograde transformation. For this reason, no hypotheses were made and this aspect of the experiments was necessarily exploratory.

6.2 Experiment 7: Unimodal trials

The aim of Experiment 7 was to investigate the perception of inverse and retrograde transformations of auditory pitch pattern structure when participants were neither instructed to compare patterns, nor informed of the types of transformation that might relate them. In each trial of the experiment, participants were presented with an auditory prime followed by an auditory target pattern. The prime was either structurally related (inverse, retrograde) or unrelated to the target. In each case, the participants' task was simply to wait until the final tone of the target stimulus and to indicate whether it was higher or lower in pitch than the preceding tone. This pitch comparison task has been used successfully in previous experiments investigating melodic expectations (Aarden, 2003; Ellis & Jones, 2009; Jones, Johnston, & Puente, 2006; Prince, Schmuckler, & Thompson, 2009).

The main hypothesis predicted that pitch comparison responses to target tones would be facilitated by related primes – responses to tones embedded in related target patterns were predicted to be faster and more accurate than responses to tones embedded in unrelated target patterns. The 1½-D hypothesis also predicted that the greatest facilitation effect would be observed when patterns were related under inverse transformation. Finally, an ISI of three different durations separated prime and target stimuli. This aspect of the experiment was exploratory, therefore no specific hypotheses were made.

6.2.1 Methods

6.2.1.1 ***Participants***

32 students from the University of Roehampton took part in Experiment 7 (female = 23, male = 9; mean age = 21.50 years, *SD* = 3.96). All had normal

hearing and normal or corrected-to-normal vision. Six participants were left-handed, and the remainder were right-handed. Nine participants reported some previous musical training (mean = 2.87 years). They all received course credit for their participation.

6.2.1.2 *Design*

A structural priming paradigm was employed in which participants were presented with two melodies, a prime followed by a target, that were separated by an ISI of varying duration. Participants compared the pitches of the final two tones of the target melody, indicating whether the final tone was higher or lower in pitch than the preceding tone. The design was 3 x 3 within-subjects, with factors target type (inverse, retrograde, unrelated) and inter-stimulus interval (ISI) (500ms, 2000ms, 4000ms). The dependent variables were correct response time (RT) (measured from the onset of the final tone of the target) and error rate (operationalised as percentage error [PE]). Unlike the previous recognition experiments reported in this thesis, the task was relatively easy and it was anticipated that the accuracy performance would be close to ceiling. For this reason RT was treated as the principal dependent variable in the present experiment (which is usual for priming experiments; Brunel, Labeye, Lesourd, & Versace, 2009; Marmel, Tillmann, & Delbé, 2010). There were 16 trials in each condition, presented in 4 blocks. The total 144 trials were randomised across blocks.

The present experiment represents an initial attempt to explore the time course of mental processing involved in the recognition of sequential patterns under inverse and retrograde transformation, and for this reason it was difficult to choose specific ISIs based on any existing theoretical or experimental evidence. As a starting point, ISIs were chosen that were multiples of 500ms so that the tones of target patterns fell on the same isochronous rhythm that had been instantiated by the prime (a 500ms inter-onset interval separated the component tones of prime and target stimuli). This step was taken as it has been shown that the identification of target pitches that follow isochronous melodic contexts can be negatively affected when they are temporally unexpected, i.e. when they do not fall on the pulse of the prior isochronous rhythmic context (Jones et al., 2006). With this restriction, a 500ms ISI was selected because it was the shortest duration available. A longer ISI of 2000ms was selected as an intermediate duration because the results at this ISI could be compared with those of the recognition experiments reported in Chapters 4 and 5 of the thesis, in which standard and target patterns were also separated by this time period. The ISI of 4000ms was selected on the basis that it was longer than the duration of prime and target stimuli (3000ms).

6.2.1.3 ***Stimuli***

6.2.1.3.1 Pattern structure

All patterns were 5 scalar-temporal relations in length (or 6 events). Relations were of equal interval size (i.e. they were isometric). All possible combinations of relation (downwards [D], same [S], upwards [U]) in a length-5 string were generated, providing a list of 243 patterns that could be transformed in three ways (inverse, retrograde, retrograde inverse). Any patterns that matched a transformation of another pattern in the list, meaning they were structurally related and therefore not *transformationally distinct*, were discarded. An additional pattern (pattern SSSSS) was discarded because it did not produce any transformational variants. Next, any pattern that began or ended with the S relation was discarded. The remaining patterns were split into two groups: one group that comprised 4 transformational variants (the original pattern and 3 others produced by inverse, retrograde and retrograde inverse transformation), and another group that comprised only 2. Those that possessed 2 transformational variants were set aside and pooled for use as unrelated targets.

16 of the 21 remaining patterns were selected as prime and related target stimuli, and one of their transformational variants was selected as the prime, under the following requirements. Constraint **A**: An equal proportion of primes ended on U or D. This ensured that, for inverse transformations, there was an equal chance of the final relation being U or D. Constraint **B**: An equal proportion of primes began on U or D. This ensured that, for retrograde transformations, there was also an equal chance of the final relation being U or D. Constraint **C**: An equal proportion of primes either began and ended with different relations or began and ended with the same relation. This ensured that, for retrograde transformations, there was an equal chance of the final relation being either the same or different to the final relation of the prime. For responses to inverse targets, the final relation was always different to the final relation of the prime, by nature of the inverse transformation (inversion of ordinal scalar relations). At this point, transformational variants of the prime that were produced by retrograde inverse transformation were discarded. The final set of prime and target stimuli used in the experiment are displayed in Table 6.1.

Table 6.1

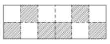
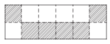
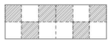
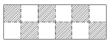
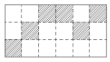
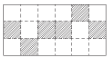
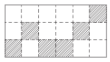
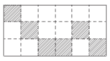
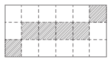
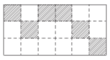
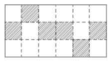
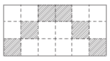
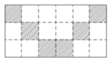
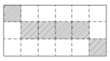
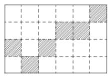
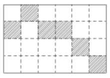
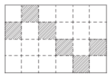
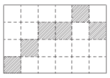
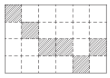
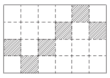
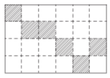
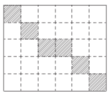
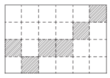
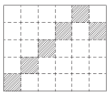
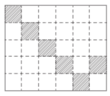
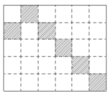
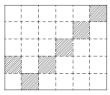
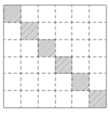
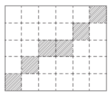
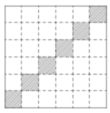
Prime and related target patterns used in Experiment 7

Prime	Related Target	
	Inverse	Retrograde
UDSSU	DUSSD	DSSUD
DDUSU	UUDSD	DSDUU
DDSSU	UUSSD	DSSUU
UUDDU	DDUUD	DUUDD
UDUUD	DUDDU	UDDUD
UUDUD	DDUDU	UDUDD
USUSD	DSDSU	USDSD
USSUD	DSSDU	UDSSD
DUDSD	UDUSU	USUDU
DSUUD	USDDU	UDDSU
DDUSD	UUDSU	USDUU
DDDUU	UUUDD	DDUUU
DSUUU	USDDD	DDDSU
USSUU	DSSDD	DDSSD
UDUUU	DUDDD	DDDUD
DDDSD	UUUSU	USUUU

Finally, unrelated targets were selected from the pool of unused patterns that had been set aside. Counting all transformational variants, there were 44 of these. It was important to ensure that these patterns did not share the same first 4 relations as any of the selected primes and their transformational variants. Though it had already been determined that all patterns were transformationally distinct, the nature of the task meant that responses to the final relation were theoretically influenced by the first 4 relations of a target pattern. Any patterns that shared the first 4 relations with those of any transformational variant of the prime (but not the 5th relation) would potentially prompt an incorrect response. Therefore, any such patterns were discarded, leaving 33 patterns for selection as unrelated targets (see Table 6.2). A final constraint was applied to the selection of unrelated targets in the experiment. Constraint **D**: An equal proportion of unrelated targets were selected that either ended on the same pitch relation or on a different pitch relation to the prime. Taken together, constraints **A**, **B**, **C** and **D** ensured there was an equal probability of any target ending on either U or D.

Table 6.2

Unrelated target patterns used in Experiment 7 (grouped by final relation)

Group A		Group B	
USSSD		DSUDU	
UDSUD		DSSSU	
DSUSD		DUSDU	
DUDUD		UUSDU	
DUSUD		UDSUU	
DDSUD		USSSU	
DUSDD		UDSDU	
UUSDD		DDSUU	
DSSSD		DUUSU	
UDSDD		UDDDU	
UUSUD		DDSDU	
DUUUD		DSDDU	
DDSDD		DUSUU	
UUUUD		DDDDU	
UDDDD		DUUUU	
DDDDD		UUSUU	
		UUUUU	

Practice stimuli were selected from discarded patterns that were transformationally distinct from any of the patterns used as related or unrelated stimuli. As the practice task was simply constructed to familiarise participants with the nature of the task (comparing the pitches of the final two target tones), it was not important whether primes and targets were related or not. All practice targets ended on U or D.

6.2.1.3.2 Pitch patterns

All pitch patterns were 6 tones in length, monophonic, isochronous and isometric. All pitches were synthesised using complex harmonic tones (triangle wave) with fundamental frequencies (f_0) taken from a 5-note equal temperament tuning. All prime stimuli began on the same tone (f_0 520.00 Hz), and a single interval on the scalar dimension equated to a change in frequency corresponding to one interval of the pitch scale. Target stimuli began on a tone with f_0 of either 640.20 Hz, or 422.37 Hz. These correspond to scale steps of one and a half intervals either above or below the first tone of the prime, and ensured that target tones had different pitches to prime tones. Each stimulus tone was 350ms in duration with linear onset and offset amplification ramps of 10ms. The inter-onset interval (IOI) between tones was 500ms and an ISI of 150ms separated tones in the same sequence. The duration of each sequence, measured from the onset of the first tone to the offset of the sixth tone was 2850ms.

All tones were generated using NCH Tone Generator version 3.02 (NCH Software), and then edited using WavePad Sound Editor Masters Edition version 5.02 (NCH Software). They were digitally recorded as .wav file type, sample size 16 bit, sample rate 44 kHz, format PCM uncompressed, mono.

6.2.1.4 ***Procedure***

Every effort was made to maintain a testing environment that was in keeping with that of the previous experiments reported in this thesis. Participants were tested individually in the same laboratory and on the same computer as before. A suitable level of instruction and training was also provided by the experimenter before the participant was left alone to complete the experimental trials.[18] The time course of an example experimental trial can be seen in Figure 6.1.

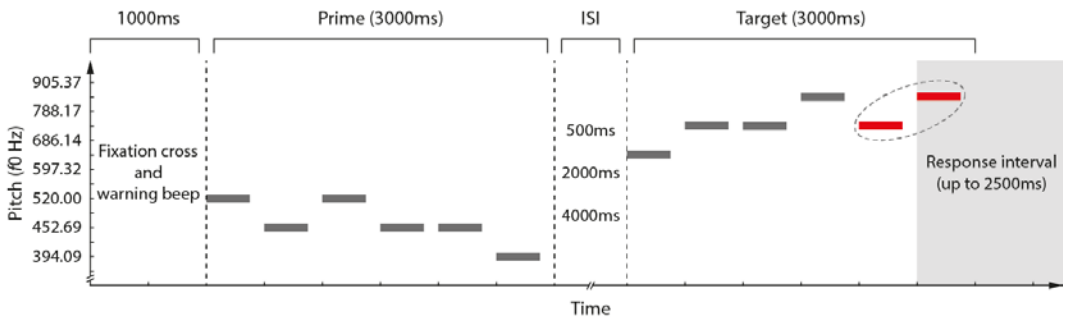


Figure 6.1. A time-pitch plot demonstrating the timeline of an experimental trial in Experiment 7. Each trial began with a warning beep. A prime melody was followed, after a short ISI, by the target melody. Here the target is a retrograde transformation of the prime (it is transposed to begin on a higher pitch than the prime – an equal proportion of targets were transposed to begin on a lower pitch). It could also be an inverse transformation or unrelated to the prime. The task was to indicate whether the final tone of the target melody was higher or lower in pitch than the tone that preceded it (highlighted in red). Response times were measured from the onset of the final tone.

On arrival participants were supplied with a briefing sheet and asked to sign a consent form. They then completed a short questionnaire collecting

[18] NB Participants were instructed and trained to compare the pitch of different tones – they were not told to compare prime and target patterns, and they were given no information regarding the transformations that might relate them.

demographic and musical background information before entering the experimental room, where they were sat in front of a computer screen for the experimental session.

In each trial participants heard two melodies, presented one after the other. The apparatus was the same as used in previous experiments. Each trial began with a warning beep (520 Hz sine wave pure tone, 200ms in duration with linear onset ramp of 3ms and offset ramp of 6ms) accompanied by a fixation point that appeared simultaneously at the centre of the screen. 1000ms after the onset of the warning beep and fixation cross the first melody was presented (the prime). After this, there was a short pause of varying duration (the ISI between the offset of the prime and the onset of the target was 500ms, 2000ms, or 4000ms). This was followed by a second melody (the target). Participants were instructed to listen to both melodies and wait for the final two tones of the second melody. The task was to indicate on a response box whether the final tone of the target was higher or lower in pitch than the preceding tone. They used the index finger of each hand to give responses on two adjacent buttons, arranged horizontally on the response box.

In 50% of the experimental sessions, 'higher' responses were assigned to the right button and 'lower' responses to the left button. In the other 50%, 'higher' responses were assigned to the left button and 'lower' responses to the right button. Participants were encouraged to respond as quickly as possible whilst maintaining accuracy. They had 2 seconds to respond once the second melody had finished, before the experiment moved on automatically to the next trial. If participants did not respond in this time they heard an alert accompanied by a message on screen prompting them to 'Please try and respond more quickly'. The alert consisted of two successive beeps each being a pure tone of 2080 Hz, 100ms in duration, 3ms onset 6ms offset, separated by an ISI of 30ms (230ms total duration). If they gave an incorrect response they heard a different alert accompanied by a message saying ‘Error’ (in red font). This alert consisted of a saw wave tone with f_0 1230 Hz, 200ms in duration. No feedback was given for correct responses. Regardless of the response (or lack thereof) there was a 1000ms pause before the next trial started. Responses were accepted once the target had started.

Before commencing the experimental trials, participants underwent some instruction and training. They were initially informed that their task would be to compare the pitches of two sequentially presented tones, and to indicate whether the second tone was higher or lower in pitch than the first. They then judged ten example tone pairs. This allowed the participant to familiarise themselves with the task, but also meant that the experimenter could advise any participants who were unfamiliar with the concept of ‘high’ and ‘low’ pitch. Examples were sampled from a list of all possible tone combinations that would be presented as the final two tones of targets in the experiment. Tone durations, IOI and ISI were the same as for the experimental stimuli. Participants responded using the response box (there was no time limit). Feedback was given on-screen including accuracy and response time. As they became more comfortable with the task, they were encouraged to reduce the response time whilst maintaining accuracy.

Once the introduction and training had finished, participants took part in 9 practice trials. Once this was finished, the experimenter left the room and the participant pushed space bar to begin the experimental trials. Every 36 trials, participants were given the opportunity to take a break before continuing by pressing spacebar. On the second break, it was compulsory to wait 2 minutes before continuing. Once the experiment was finished, the experimenter recorded comments on the participants' experience of the task. In particular, they were asked if they noticed anything about the relationship between the prime and target melody that might have influenced their responses. The entire experimental session took approximately 60 minutes to complete.

6.2.2 Results

Data from 10 participants had to be excluded from the analysis because they failed to achieve accuracy scores of 70% correct or above. This was surprising and contradicted the previous assertion that the task would be easy. Inspection of the individuals' data suggested that three of these participants did not engage with the task and frequently responded before the onset of the final tone of the target. It is possible that the remaining participants had relatively poor pitch perception abilities. The present experiment did not include a pitch perception screening test (the experimental training did include 10 example tone pairs and 10 practice trials, but they were not excluded from taking part in the experiment if they performed poorly in these tasks).

6.2.2.1 ***RT data***

6.2.2.1.1 Target type

Overall mean RT was 771.78ms (SD = 180.57). Figure 6.2 displays the mean RT for the target type conditions, collapsed across ISI conditions. Responses were fastest in the retrograde condition and slowest in the unrelated condition. The transformation priming hypothesis predicted that performance in the pitch comparison task would be facilitated by related primes. To test this hypothesis an initial one-way ANOVA examined the within-subjects effects of target type (inverse, retrograde, unrelated) on log-transformed RT data. The analysis revealed a significant main effect of target type, $F(2,42) = 5.67$, $MSE < .01$, $p = .007$, $\eta_p^2 = .21$, indicating the structural relationship between the prime and target had an influence on the speed of pitch comparison responses.

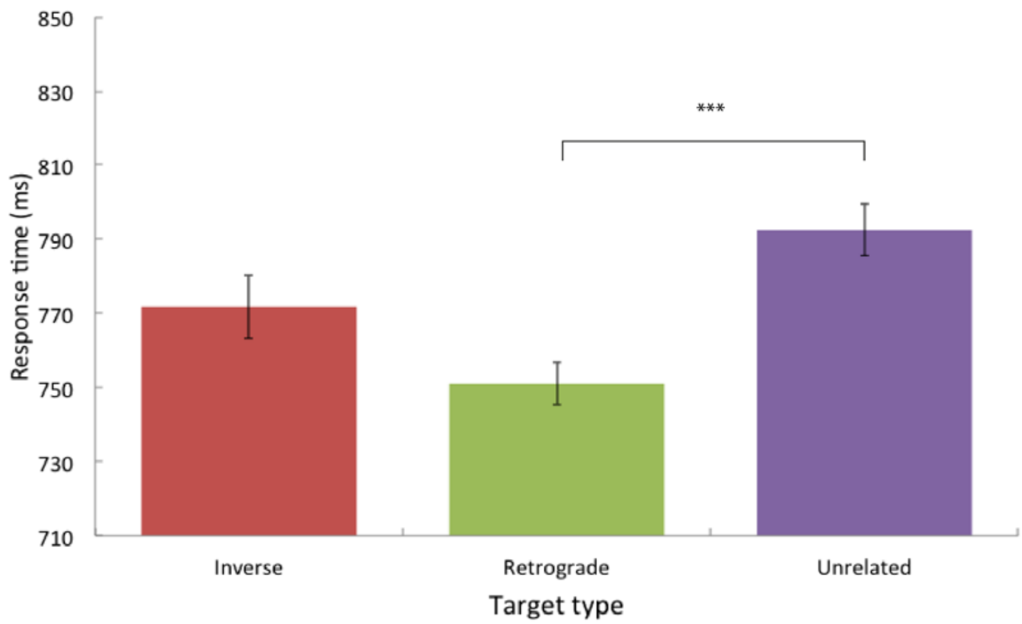


Figure 6.2. Experiment 7: Mean RT for inverse, retrograde and unrelated target conditions. Significance values for simple effects were obtained from the analysis on log-transformed data ($^{ms}p < .1$ [marginally significant], $*p \leq .05$, $**p < .01$, $***p < .001$). Error bars indicate the standard error of the mean (SEM ± 1).

Pairwise comparisons were run to examine the simple effects of target type. As these tests were planned no correction was applied to the alpha level. Comparisons revealed the mean difference between inverse and unrelated conditions failed to reach significance ($MD = 20.72$, $SE = 14.54$; $p = .133$) but the mean difference between retrograde and unrelated conditions was highly significant (*MD* = 41.57, *SE* = 9.39; p = .001). Although the facilitation effect of primes in the inverse condition failed to reach formal significance, these results provided some support for the transformation priming hypothesis.

The 1½-D hypothesis predicted that the facilitation effect would be greatest when prime and target patterns were related under inverse transformation. However, as can be seen in Figure 6.2, mean RT was actually fastest in the retrograde condition, though the mean difference between conditions was non-significant (*MD* = 20.84, *SE* = 12.88; p = .109). This result failed to support the 1½-D hypothesis.

6.2.2.1.2 Target type by ISI

Further analysis was carried out to explore the main effect of ISI, the interaction between ISI and target types, and the simple effects of target type and ISI on log-transformed RT data. A 3 x 3 within-subjects ANOVA was run on target type (inverse, retrograde, unrelated) and ISI (500ms, 2000ms, 4000ms). The main effect of target type has already been reported above. The main effect of ISI, $F(2,42) = 1.34$, $MSE = .01$, $p = .274$, $\eta_p^2 = .06$, and the interaction between ISI and target type, $F(4,84) = 0.81$, $MSE = .01$, $p = .522$, $\eta_p^2 = .04$, were both non-significant. No further analysis was carried out.

6.2.2.2 ***Error data***

6.2.2.2.1 Target type

Overall mean PE was 14.74 (*SD* = 8.77), which was quite high for an experiment that treated RT as the primary dependent variable. This may pose some difficulties for the interpretation of the analyses based on RT data, reported above. Figure 6.3 displays the mean PE for the target type conditions, collapsed across ISI conditions. The largest amount of error was made in the inverse condition, and the least amount of error in the retrograde condition. The same analysis that was carried out on log-transformed RT data was carried out on arcsine transformed PE data. An initial one-way ANOVA revealed a significant within-subjects effect of target type (inverse, retrograde, unrelated), $F(2,42) = 6.93$, $MSE = .01$, $p = .003$, $\eta_p^2 = .25$. This result was in agreement with the analysis on RT data, indicating that the structural relationship between prime and target patterns influenced performance accuracy in the pitch comparison task.

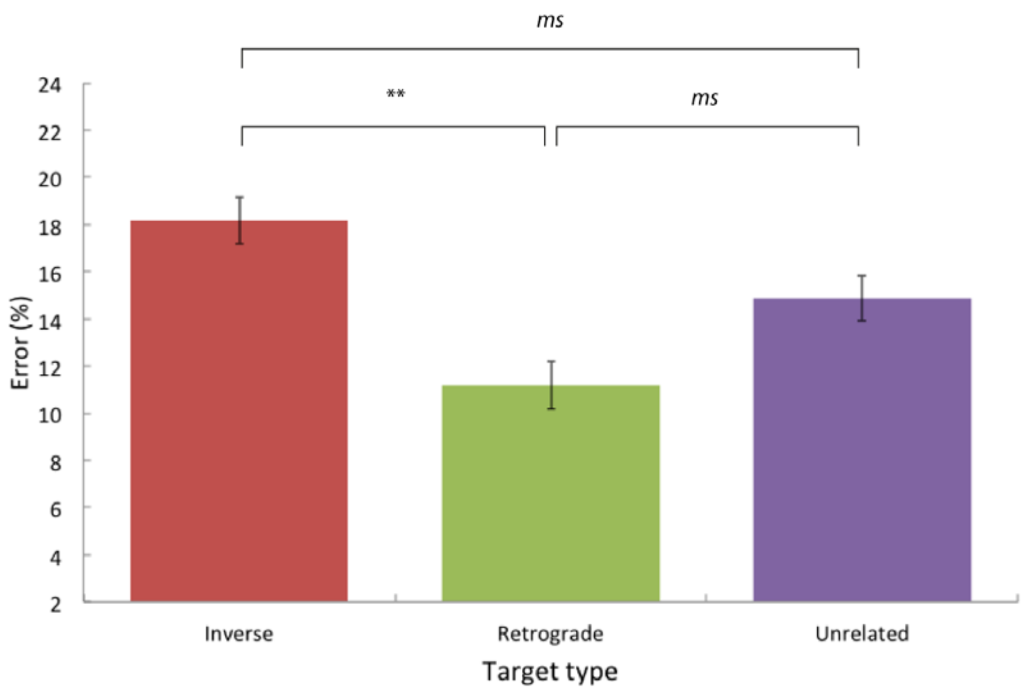


Figure 6.3. Experiment 7: Mean PE for inverse, retrograde and unrelated target conditions. Significance values for simple effects were obtained from the analysis on arcsine-transformed data ($^{ms}p < .1$ [marginally significant], $*p \leq .05$, $**p < .01$, $***p < .001$). Error bars indicate the standard error of the mean (SEM ± 1).

As predicted by the main transformation priming hypothesis, PE was lower in the retrograde condition than in the unrelated condition, but the mean difference merely approached significance (MD = 3.69, SE = 1.71; p = .057). PE in the inverse condition was actually higher than in the unrelated condition, which suggested that responses may have been impaired by primes in this condition. However, the mean difference also only approached significance (MD = 3.31, SE = 1.69; p = .095). The 1½-D hypothesis predicted that the facilitation effect would be greatest in the inverse condition. As neither retrograde nor inverse conditions facilitated responses there was little need to perform any further analysis to address this hypothesis. Nevertheless, there was a significant mean difference between the inverse and retrograde conditions (MD = 7.01, SE = 1.74; p = .002), with less error being made in the retrograde condition. This result failed to support the hypothesis.

6.2.2.2.2 Target type by ISI

An exploratory analysis was carried out to investigate the pattern of results over the different ISIs. A 3 x 3 within-subjects ANOVA was run on target type (inverse, retrograde, unrelated) and ISI (500ms, 2000ms, 4000ms). Both the main effect of ISI, $F(2,42) = 0.49$, $MSE = .02$, $p = .619$, $\eta_p^2 = .02$, and the interaction between ISI and target type were non-significant, $F(4,84) = 1.08$, $MSE = .02$, $p = .373$, $\eta_p^2 = .05$. No further analysis was carried out.

6.2.2.3 ***Music training analysis***

An additional exploratory analysis was carried out to see if participants' music training experience had any effect on their performance in the experiment (see Appendix IV for the full results – only significant findings are reported here). In order to perform this analysis, participants were allocated to a 'some training' or a 'no training' group, according to their self-reports. Only 6 participants reported that they had previously received some music training. 16 participants reported that they had received no previous music training. Firstly, a Pearson's correlation was conducted to compare the amount of music training reported by participants in the some training group with performance (RT and PE data) in the target type conditions (inverse, retrograde, unrelated). There were no significant correlations.

Secondly, the effects of music training were explored by repeating the analysis of target type on RT (log-transformed) and PE (arcsine transformed) data, collapsed across ISI conditions, and including music training as an additional between-subjects factor. Participants were allocated to one of two groups according to their self-reports. Only 6 participants reported that they had previously received some music training, and were allocated to a 'some training' group. 16 participants reported that they had received no previous music training and were allocated to a 'no training' group. 2 x 3 mixed ANOVAs with target type as the within-subjects factor (inverse, retrograde, unrelated) and music training as the between-subjects factor (no training, some training) were run on RT (log-transformed) and PE (arcsine transformed) data. The analysis failed to reveal any significant results.

6.2.3 Discussion

The analysis on PE data failed to reveal any formally significant facilitation effects of transformation. However, the analysis on RT data provided some support for the main transformation priming hypothesis – compared to unrelated patterns, responses were significantly faster when patterns were related under retrograde transformation, but not when patterns were related under inverse transformation. This suggests participants perceived retrograde transformations, which enabled them to generate specific (and correct) expectations about how related target patterns would unfold. In turn, expected tones were perceived more effectively, enabling faster pitch comparisons. It is possible that expected tones were perceived more effectively because attentional resources were directed towards the appropriate pitch range (Dowling et al., 1987; Greenberg & Larkin, 1968; Howard et al., 1984; Jones, 1990; Tanner & Norman, 1954). The 1½-D hypothesis, on the other hand, was not supported. Although there was no significant difference between the two conditions, only retrograde transformations facilitated responses.

The results contribute to the previous recognition experiments reported in this thesis, and show that retrograde transformations of auditory patterns are perceived when participants have been given no explicit instruction or training. The absence of any facilitation effects of inverse transformation was surprising, considering that not only had they been recognised in these previous experiments, but that they had been recognised more successfully than retrograde transformations (Experiment 6).

The results also contribute to previous research that has demonstrated the influence of pattern-based expectancies on indirect perceptual tasks (Dowling et al., 1987; Howard et al., 1984; Johnston & Jones, 2006). However, whilst previous experiments had demonstrated that expectancies may be based on the abstraction of simple periodic patterns (Johnston & Jones, 2006), the present research goes further by demonstrating that expectancies may be based on the processing of retrograde transformations. More broadly, the results of the present experiment support classic theories of serial pattern learning (Jones, 1976a; Restle & Brown, 1970; Restle, 1970) by demonstrating that the relationships between patterns under retrograde transformation are abstracted and used in an experimental setting when participants have not been explicitly instructed to attend to the pattern as a whole.

An additional aim of the present experiment was to explore the time course of processes involved in the recognition of patterns under inverse and retrograde transformation. The influence of structural relations on performance in the pitch comparison task was analysed across three different ISIs that separated the prime and target patterns, revealing two important findings. Firstly, the influence of structurally related patterns on performance in the pitch comparison (facilitation or impairment) task did not change over time. This suggests that participants were able to maintain conscious expectations over a reasonable period of time. Previous research in language processing that manipulated ISI in priming experiments demonstrated that facilitation effects associated with expectations usually diminish as ISI increases (e.g. Carter et al., 2011), presumably caused by the decay of a memory trace for the prime stimulus. The time period of the longest ISI in the present experiment (4000ms) was presumably not sufficiently long to reveal the diminishing effects of prime structure on responses. Future experiments could test this possibility by incorporating longer ISIs. Secondly, the influence of structural relatedness was evident at the shortest ISI, indicating that the mental transformation required to perceive that the target was related to the prime had already been performed within 500ms of the prime stimulus being presented. The implication is that mental transformations of sequential pattern structure are applied on-line to any incoming sensory input, as opposed to being applied to a pattern as a whole once it has been encoded.

In conclusion, Experiment 7 provided some support for the main transformation priming hypothesis, demonstrating that retrograde transformations of auditory pitch patterns were perceived and therefore facilitated responses in a pitch comparison task. The absence of any significant facilitation effects of inverse transformation meant that the 1½-D hypothesis was unsupported. Further exploratory analysis of the effects of ISI on performance failed to reveal any significant effects.

6.3 Experiment 8: Cross-modal trials

The aim of Experiment 8 was to investigate the perception of inverse and retrograde transformations of sequential pattern structure in cross-modal trials when participants were neither instructed to compare patterns, nor informed of the types of transformation that might relate them. In each trial, participants were presented with an auditory prime followed by a structurally analogous visual target pattern. Visual stimuli comprised sequences of objects presented at different vertical heights. Both auditory and visual stimuli corresponded to representations in a 1½-D supramodal pattern space. The prime was either structurally related (inverse, retrograde) or unrelated to the target. The task was to wait until the final object of the visual target and to indicate whether it was higher or lower (in vertical space) than the preceding object.

The main transformation priming hypothesis predicted that perceptual responses would be facilitated by related primes. More specifically, the 1½-D hypothesis predicted that the facilitation effect of related primes would be greatest when they were related to target patterns under inverse transformation, compared to retrograde transformation. Finally, the effects of three different inter-stimulus intervals (ISI) were explored. No specific hypotheses were made regarding the effects of ISI in the different conditions.

6.3.1 Methods

6.3.1.1 *Participants*

30 participants from the University of Roehampton took part in Experiment 8 (female = 24, male = 6; mean age = 21.20 years, *SD* = 4.77). All reported normal hearing and normal or corrected-to-normal vision. Twenty-seven were right-handed, and the remainder were left-handed. Five participants reported some previous music training (mean = 5.20 years). They all received course credit for their participation.

6.3.1.2 *Stimuli*

The same pattern structures that were used in Experiment 7 were used in Experiment 8. Auditory stimuli were produced in the same way as they were in Experiment 7. All auditory primes began with a 520.00 Hz tone. Visual stimuli were sequential spatial patterns, and were produced to be analogous with auditory stimuli, in the same way as in Experiment 6. They consisted of a sequence of six black bar segments presented at different heights on the vertical axis. The first segment of each stimulus was presented at the centre of the screen. Each scalar-temporal relation of the pattern structures was equivalent to two successive black bar segments. A U relation was equivalent to an upwards displacement on the vertical axis of 1.35° of visual angle (as determined by Experiment 1), a D relation was equivalent to a downwards displacement on the vertical axis of the same distance, and a S relation was equivalent to no displacement on the vertical axis. Each segment measured 2.23 by 0.45° visual angle. The IOI was 500ms and the duration of each segment was 350ms. Thus, there was an ISI of 150ms between the offset of one segment and the onset of another. The duration of a stimulus from the onset of the first segment to the offset of the final segment was 2850ms.

Visual stimuli were produced using Adobe Illustrator and were animated in Final Cut Pro. Stimuli were saved as WMV files and presented using E-Prime software.

6.3.1.3 *Design and procedure*

The experimental design was identical to Experiment 7, replacing auditory targets with visual targets. The procedure was also identical, except for the following exceptions. In each trial an auditory prime was followed by a visual target. Participants were instructed to respond to the final object of the visual target, indicating whether it was higher or lower than the preceding object. The appropriate changes were made to the pre-experiment instructions and training. As the task of identifying whether a visual segment was spatially above or below another object was more readily understood by participants, examples of visual stimuli were given, but there was no training phase equivalent to that provided in Experiment 7 (in which participants indicated whether the second of two tones was higher or lower in pitch than the first).

6.3.2 Results

The same analysis that was run on the data from Experiment 7 was performed on the data collected from the present experiment. Data from 7 participants were excluded from the analysis because they failed to achieve accuracy scores of 70% correct or above. Inspection of these participants' data suggested that their low accuracy scores were a result of not engaging with the task – a large proportion of responses were made before the final segment of the target pattern was presented, resulting in a large proportion of responses that were deleted from the data set.

6.3.2.1 ***RT data***

6.3.2.1.1 Target type

An initial one-way ANOVA was run to examine the within-subjects effects of target type (inverse, retrograde, unrelated) on log-transformed RT data. Overall mean RT was 805.02ms (*SD* = 203.63), which was approximately 33ms slower than in Experiment 7. This is in agreement with previous literature that has generally shown that reaction times to visual stimuli are slower than reaction times to auditory stimuli (e.g. Shelton & Kumar, 2010), something that has been known since the classic research carried out by Francis Galton in the 19th Century (Johnson et al., 1985). Figure 6.4 displays the mean RT for the target type conditions, collapsed across ISI conditions. Mean RT was fastest in the retrograde condition and slowest in the inverse condition. The main effect of target type failed to reach significance, $F(2,44) = 1.04$, $MSE < .01$, $p = .363$, $\eta_p^2 = .05$.

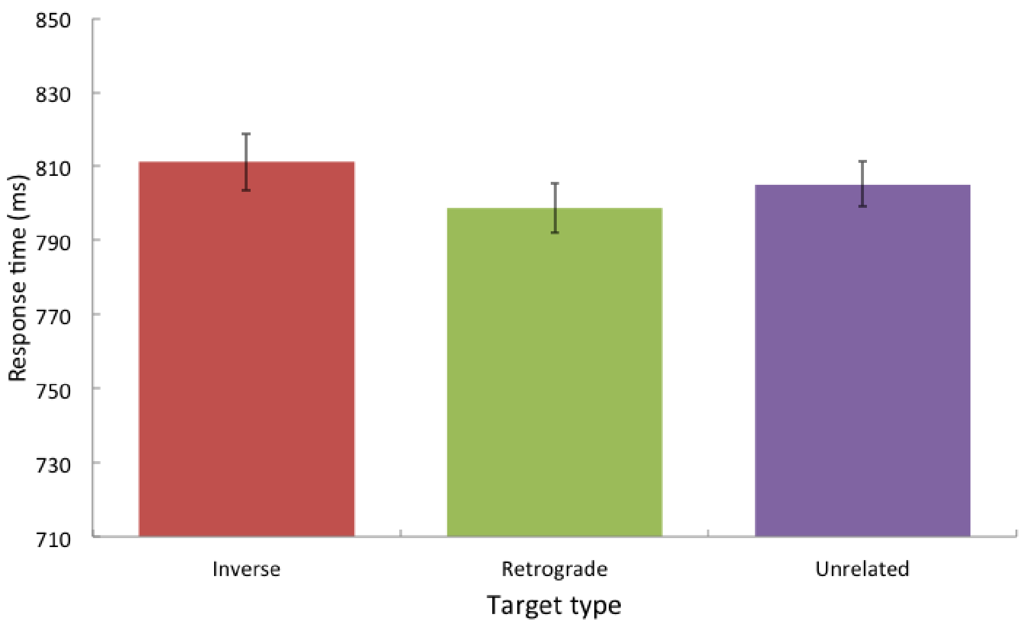


Figure 6.4. Experiment 8: Mean RT for inverse, retrograde and unrelated target conditions. Error bars indicate the standard error of the mean (SEM ± 1).

The main transformation priming hypothesis predicted that responses would be facilitated in related conditions compared to the unrelated condition. Therefore, despite the failure to find a significant main effect of target type, pairwise comparisons were run to examine the simple effects of target type. When compared to the unrelated condition, mean RT was faster in the retrograde condition but slower in the inverse condition. This suggested that responses were facilitated in the retrograde condition and impaired in the inverse condition (as was revealed by the analysis on PE data in Experiment 7). However, both mean differences were non-significant (inverse [MD = 6.15, SE = 10.66; p = .162], retrograde [MD = 6.37, SE = 10.90; p = .856]). These results failed to support the transformation priming hypothesis.

A more specific 1½-D hypothesis predicted that the facilitation effect would be greatest in the inverse condition. Mean RT in the inverse condition was actually slower than in the retrograde condition, though the mean difference was non-significant (*MD* = 12.52, *SE* = 10.42; $p = .245$). This result failed to support the 1½-D hypothesis.

6.3.2.1.2 Target type by ISI

Further analysis was carried out to explore the effects of ISI on log-transformed RT data in each target condition. A 3 x 3 within-subjects ANOVA was run on target type (inverse, retrograde, unrelated) and ISI (500ms, 2000ms, 4000ms). The main effect of target type has already been reported above. The main effect of ISI was approaching significance, $F(2,44) = 2.60$, $MSE = .01$, $p = .085$, $\eta_p^2 = .11$. No further analysis was carried out.

6.3.2.2 ***Error data***

6.3.2.2.1 Target type

The same analysis that was carried out on log-transformed RT data was carried out on arcsine transformed PE data. Overall mean PE was 7.16 (*SD* = 5.85), which was markedly less than was made by participants in Experiment 7 (*MD* = 8.32). Figure 6.5 displays the mean PE for the target type conditions, collapsed across ISI conditions. Mean PE was lowest in the retrograde condition and highest in the inverse condition. An initial one-way ANOVA revealed a significant within-subjects effect of target type (inverse, retrograde, unrelated), $F(2,44) = 6.14$, $MSE = .01$, $p = .004$, $\eta_p^2 = .22$. This result was in contrast to the result of the analysis on RT data, which failed to find a significant main effect of target type, and indicated that the structural relationship between prime and target patterns influenced performance accuracy in the pitch comparison task.

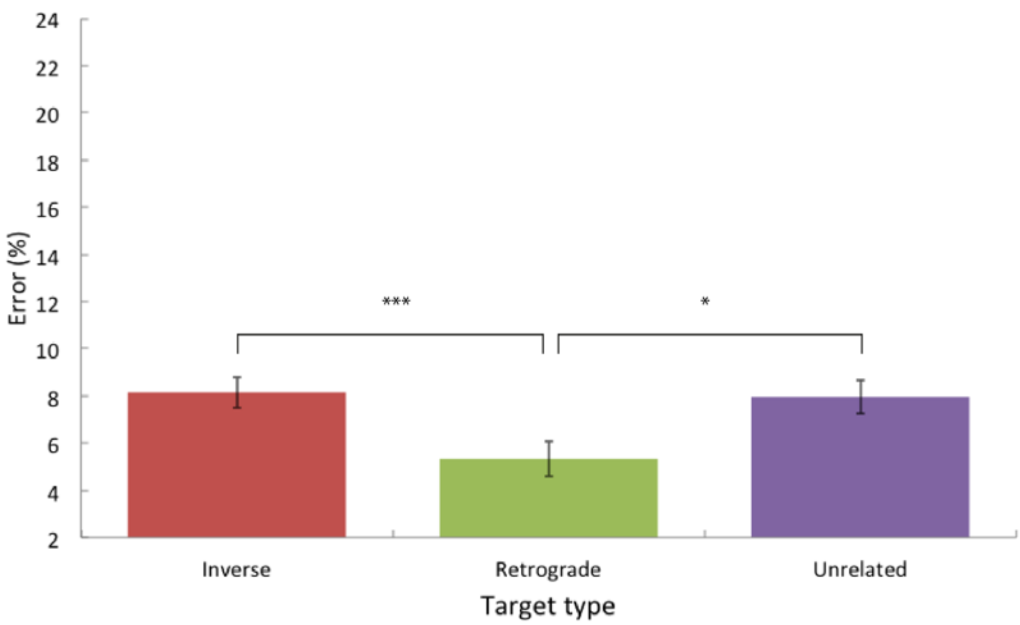


Figure 6.5. Experiment 8: Mean PE for inverse, retrograde and unrelated target conditions. Significance values for simple effects were obtained from the analysis on arcsine-transformed data ($^{ms}p < .1$ [marginally significant], $*p \leq .05$, $**p < .01$, $***p < .001$). Error bars indicate the standard error of the mean (SEM ± 1).

Pairwise comparisons were run to examine the main transformation priming hypothesis, which predicted that performance would be facilitated in related conditions compared to the unrelated condition. These revealed a significant mean difference between unrelated and retrograde conditions (MD = 2.63, SE = 1.36; p = .021), with less error being made in the retrograde condition, supporting the hypothesis. Mean PE was actually slightly higher in the inverse condition compared to the unrelated condition, suggesting that performance might have been impaired in the inverse condition. However, the mean difference was non-significant (MD = 0.18, SE = 1.19; p = .742). An additional comparison was run on retrograde and inverse conditions to examine the 1½-D hypothesis, which predicted that the facilitation effect would be greatest in the inverse condition. Mean PE was actually higher in the inverse condition, and the mean difference was found to be highly significant (*MD* = 2.81, *SE* = 1.05; $p = .001$). This result did not support the 1½-D hypothesis.

6.3.2.2.2 Transformation by ISI

Further analysis was carried out to explore the effects of ISI on performance. A 3 x 3 within-subjects ANOVA was run on target type (inverse, retrograde, unrelated) and ISI (500ms, 2000ms, 4000ms), and revealed a highly significant main effect of ISI, $F(2,44) = 7.92$, $MSE = .02$, $p = .001$, $\eta_p^2 = .27$. The interaction between target type and ISI was non-significant, $F(4,88) = 0.48$, $MSE = .02$, $p = .747$, $\eta_p^2 = .02$.

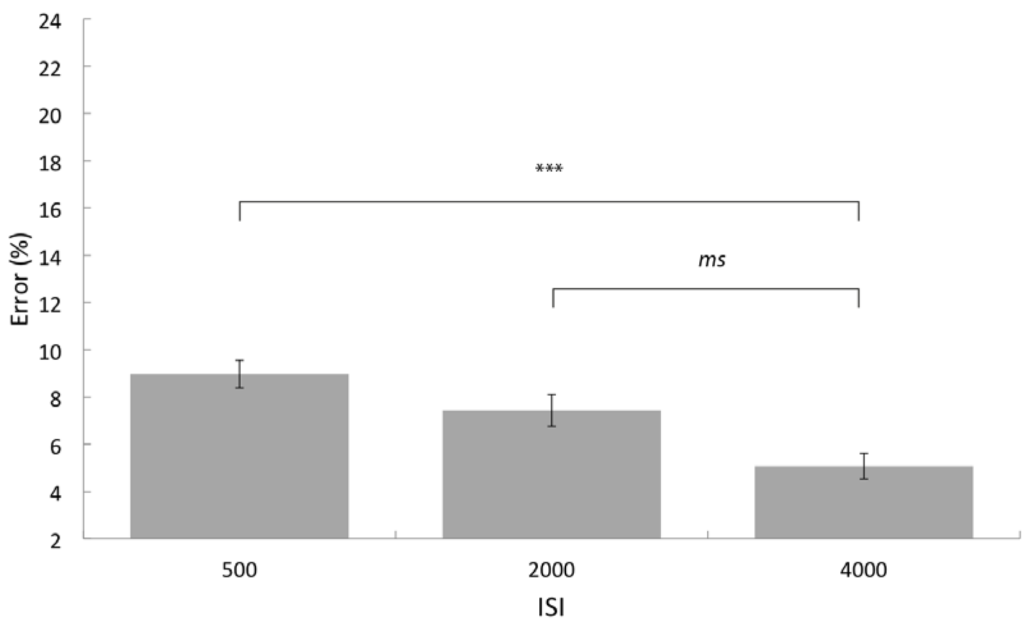


Figure 6.6. Experiment 8: Mean PE in ISI conditions ($^{ms}p < .1$ [marginally significant], $^{*}p \leq .05$, $^{**}p < .01$, $^{***}p < .001$). Error bars indicate the standard error of the mean (SEM ± 1).

In order to explore the significant main effect of ISI further, pairwise comparisons (Bonferroni-corrected) were run to examine the simple effects of ISI.

As can be seen in Figure 6.6, mean PE decreased with increasing ISI. The mean difference between 500ms ISI and 2000ms ISI was non-significant (*MD* = 1.54, *SE* = 1.14; *p* = .550). However, the mean difference between 500ms ISI and 4000ms ISI was highly significant (*MD* = 3.90, *SE* = 0.91; $p < .001$), and the mean difference between 2000ms ISI and 4000ms ISI was approaching significance (*MD* = 2.36, *SE* = 1.08; *p* = .090).

These results suggest that auditory primes impaired performance in the spatial comparison task, irrespective of their structural relationship to visual targets. Furthermore, the impairment effect decreased as the temporal proximity between primes and targets increased. Thus, it appears that two processes occurred in parallel – one that involved the perception of retrograde transformations and facilitated responses compared to unrelated (and inverse) conditions, and another where auditory stimuli in general impaired responses.

6.3.2.3 ***Music training analysis***

An exploratory analysis of the effects of music training was carried out (see Appendix IV). Only 4 participants self-reported some previous music training, whilst 19 self-reported no previous music training. Firstly, a Pearson's correlation analysis was run on the amount of music training and performance in target conditions, but revealed no significant results.

Secondly, two 2 x 3 mixed ANOVAs were run on RT and PE data, with the within-subjects factor target type (inverse, retrograde, unrelated) and the between-subjects factor music training (some training, no training). The analysis on RT data revealed a non-significant main effect of music training and a non-significant interaction. The analysis on PE data revealed a significant main effect of music training, $F(1,21) = 7.34$, $MSE = .03$, $p = .013$, $\eta_p^2 = .26$. Participants with some training ($M = 14.58$, $SD = 7.80$) made more errors in the spatial comparison task than participants with no music training ($M = 5.59$, $SD = 4.10$). This result was surprising because in the previous experiments reported in this thesis, music training has been associated with more effective perception of transformed patterns. There were no further significant results.

As noted in previous experiments, the analysis on music training was carried out post-hoc and for this reason the sample sizes in 'some training' and 'no training' groups were often unbalanced. This was particularly true in the present experiment, as only 4 participants were allocated to the some training group. Therefore, little validity can be ascribed to the above findings.

6.3.3 Discussion

The analysis on RT data failed to reveal any significant facilitation effects of transformation, so the following discussion will focus on the analysis on PE data. The key finding in Experiment 8 was that when auditory primes were related to visual targets under retrograde transformation, spatial comparison responses were significantly more accurate. This facilitation effect indicated that the structural relationship between auditory prime and visual target patterns was perceived, which enabled participants to generate correct expectations for how the target would unfold. In turn, events that were expected were easier to compare, possibly due to expectations guiding attentional focus in visual space (Posner et al., 1978, 1980; Posner & Snyder, 1975a, 1975b).

This result provided some support for the main transformation priming hypothesis and is consistent with the notion that auditory and visual information share suprmaodal structural representations and processes – the perceptual system apparently treated auditory and visual structural information equivalently. According to the SPS framework, both auditory and visual stimuli used in the present experiment corresponded to an equivalent representation in supramodal pattern space (1½-D). It is possible that the structural relationship between retrograde-transformed prime and target patterns required an inversion of ordinal relations on the temporal dimension. It would be interesting to see if the same facilitation effect persists when visual stimuli are presented that do not correspond to an equivalent supramodal pattern space, such as the 2½-D pattern space investigated in Chapter 4.

The facilitation effect was only observed when prime and target patterns were related under retrograde transformation. On the one hand, this finding failed to support the 1½-D hypothesis, which predicted that the facilitation effect would be greatest when patterns were related under inverse transformation. However, more than this, it suggests that inverse transformations were not processed at all. This result was surprising, as it has been shown in previous cross-modal recognition experiments (reported earlier in this thesis) that both retrograde and inverse transformations can be recognised in cross-modal trials (Experiments 3 and 5). The implication is that, although inverse transformations of supramodal structural information can be processed under explicit instruction, they are not processed automatically.

An additional aim of the present experiment was to explore the time courses of processes involved in the perception of structural transformations, by examining the pattern of responses across three different ISIs (500ms, 2000ms, 4000ms). The facilitation effect of retrograde transformations was evident at the shortest ISI, indicating that the necessary mental transformation had been performed within 500ms of offset of the prime, which was also the case in Experiment 7. However, whilst no effect of ISI was observed in Experiment 7, in Experiment 8 a general trend was observed whereby performance improved as ISI increased. This cannot be attributed to an increase in the facilitation effect of structurally related primes over time, as the improvement in performance occurred regardless of whether the prime and target were structurally related or not. Instead, the temporal proximity of auditory primes appeared to have a detrimental effect on the accuracy of responses to visual spatial comparisons that was independent from the processing of structural relatedness. This finding is puzzling, and indicates that visual perception was negatively affected either by auditory information in general, or by auditory structural information specifically. To examine this further, the experiment would need to be repeated with additional non-structural primes, such as a noise prime or an absent prime. When preceded by a noise prime, no effect of ISI would confirm that structural processes interfere with visual perception. On the other hand, an effect of ISI would indicate that auditory information in general interferes with visual perception, and this could be confirmed if there is no effect of ISI when an auditory prime is absent.

In conclusion, the results of Experiment 8 demonstrated that performance in a visual spatial comparison task was facilitated when visual targets were structurally related to auditory primes under retrograde transformations. This was interpreted as evidence to show that auditory and visual stimuli are processed by a shared mechanism that is sensitive to supramodal structural information. However, the results failed to provide evidence for the supramodal processing of inverse transformations. The results of Experiment 8 also demonstrated that temporal proximity of auditory and visual stimuli influenced responses – as the ISI between auditory primes and visual targets increased, performance improved in related and unrelated conditions. This suggests that either auditory information in general or auditory structural information may have interfered with perception in the visual domain.

6.4 General discussion

The general aim of Experiments 7 and 8 was to investigate the perception of inverse and retrograde transformations of structural information. In order to do this, a structural priming paradigm was adopted in which participants compared events that were embedded in a target pattern. The target pattern was preceded by a structurally related (inverse, retrograde) or unrelated prime pattern. Importantly, this paradigm allowed the study of structural processing when participants were not explicitly instructed to compare patterns and not given any prior training regarding the types of transformation that might relate them.

In Experiment 7, prime and targets were both auditory pitch patterns. Thus, the task required participants to compare the pitches of two tones. In Experiment 8, a visual target was preceded by an auditory prime. Thus, the task required participants to compare the spatial positions of two objects. Informed by the SPS framework, two hypotheses were examined: the transformation priming hypothesis, which predicted that performance in the comparison task would be facilitated when target patterns were related to primes in unimodal (Experiment 7) and cross-modal (Experiment 8) trials. A more specific hypothesis, called the 1½-D hypothesis, predicted that the facilitation effect of structural relatedness would be greatest for inverse transformations.

Table 6.3

Chapter 6: Summary of experiments, hypotheses and results

Experiment	Modality	Hypothesis tested	Result
7	A	Transformation priming	Some support
		1½-D	Not supported
8	AV	Transformation priming	Some support
		1½-D	Not supported

Note. A = auditory prime and target; AV = auditory prime, visual target

A summary of the experiments, hypotheses and results can be seen in Table 6.3. The main finding was that the transformation priming hypothesis was supported in both unimodal and cross-modal experiments, but only when patterns were related under retrograde transformation. Specifically, when auditory primes were structurally related to target patterns, performance in a perceptual comparison task was facilitated whether targets were auditory (Experiment 7) or visual (Experiment 8). This indicated that participants were able to generate correct expectations about how targets would unfold, based on the perception of a supramodal structural relationship between patterns.

The facilitation effect of retrograde transformations in unimodal and cross-modal trials is in agreement with the notion of supramodal pattern space, proposed by the SPS framework (see Chapter 2). According to the SPS framework, the facilitation effect may reflect equivalent representations of auditory and visual stimuli in a 1½-D supramodal pattern space. The perception of a structural relationship between patterns represented in this pattern space may involve a supramodal mechanism that inverts ordinal relations on supramodal dimensions – in the case of retrograde transformation, ordinal relations need to be inverted on the temporal dimension. The analysis on the effects of ISI would suggest that, in both experiments, the required mental transformation had been completed within 500ms of the offset of the prime pattern. From this evidence, it may be assumed that an inversion of ordinal relations is applied to structural information as it is abstracted from the incoming sensory input (rather than being applied to abstracted structural information on completion of the sensory input).

As discussed in the introduction (Section 6.1), previous experiments have demonstrated that melodic transformations may be recognised implicitly (Dienes & Longuet-Higgins, 2004). In the present experiment, participants were not given any training regarding the types of transformation that had been applied to patterns and were not explicitly instructed to compare prime and target patterns. It is possible that the perception of retrograde transformations in the present research was therefore based on implicit recognition. However, this claim is made with some caution. Having completed the experiments, all participants were asked to comment on their experience of the task, and in particular they were asked whether they noticed anything about the relationship between prime and target patterns that may have influenced their responses. Whilst the majority expressed surprise to learn that some of the targets were related to primes in some way, others indicated that they noticed a relationship but were unable to explain what the relationship was. A few participants, on the other hand, were not only aware that some of the patterns were related, but were also able to describe how they were related.

The transformation priming hypothesis was not fully supported, because inverse transformations did not facilitate responses in either experiment. The absence of any facilitation effects of inverse transformations meant that the 1½-D hypothesis was unsupported. The 1½-D hypothesis was based on one of the assumptions of the SPS framework – that the perception of structural relationships described by inverse transformation would require an inversion of ordinal relations on a scalar dimension, which is easier to process than inversions on the temporal dimension, due to the latter dimension's inherent directionality. It is not clear why inverse transformations would not be perceived when retrograde transformations are, especially as previous experiments reported in this thesis appear to show that inverse transformations are processed more effectively. However, it must be concluded from this evidence that, although pattern relationships described by inverse transformation can be perceived under explicit instruction, the mental processes required (i.e. an inversion of ordinal relations on a scalar dimension) are not necessarily utilised automatically.

It should be noted that there were some limitations concerning the task employed in the present experiments. Participants were required to perform different perceptual tasks in each experiment – in Experiment 7 the task was to compare the pitches of tones, whereas in Experiment 8 the task was to compare the spatial positions of visual objects. The design of Experiment 8 (a visual target preceded by an auditory prime) was chosen over the other option (an auditory target preceded by a visual prime) because it was anticipated that it would be difficult to ensure that participants would watch a visual prime when the task was to compare the pitches of two tones. Nevertheless, comparing the pitches of tones was apparently much more difficult for participants than comparing the spatial positions of visual objects (Experiment 7: mean PE = 14.74; Experiment 8: mean PE = 7.16). This may have had something to do with low levels of music training in the population samples examined – in Experiment 7 only 27% of the sample reported some previous music training, and in Experiment 8 this proportion was even lower, with only 17%. A more complete investigation of potential automatic supramodal processes would include a unimodal experiment in which prime and target patterns are both visual, and a cross-modal experiment in which an auditory target is preceded by a visual prime (provided a suitable solution can be found to ensure that participants watch the visual prime).

In conclusion, the findings reported in this chapter provide some evidence for the supramodal processing of retrograde transformations of structural information. Although this finding is in agreement with the SPS framework's general notion of a supramodal pattern space, the more specific hypothesis – that an inversion of ordinal relations on the temporal dimension is processed less effectively than an inversion of ordinal relations on the scalar dimension – was not supported. It is possible that inverse transformations of supramodal structural information are not processed automatically.

Chapter 7: General discussion and conclusions

1.1 Summary and general analysis

The experiments described in this thesis have been concerned with the possibility that auditory pitch patterns share, at some level, supramodal structural representations and processes with visuo-spatial patterns.

The motivation for the research was provided by a number of areas of psychological research that have been discussed in some detail in Chapter 1, the main strands of which will be briefly recapitulated here. Firstly, strikingly similar (if not the same) principles appear to govern pattern perception in both the auditory and the visual domains. Secondly, there is growing evidence for the spatial representation of psychological dimensions such as auditory pitch and time. Thirdly, neuropsychological studies have shown that auditory pitch patterns and visuo-spatial patterns may be processed in shared higher-order anatomical areas of the cortex. Specifically, areas of the posterior parietal cortex have been associated with the processing of melodic transformations and visuo-spatial transformations.

The main aim of the research reported here was to explore possible supramodal processes more thoroughly and in more detail than has been achieved before now, by means of behavioural experimentation. To this end, a theoretical framework has been proposed in Chapter 2 that conceives of a supramodal pattern space (SPS). According to the SPS framework, structural information, abstracted from sensory information, can be represented on one or a combination of qualitatively distinct supramodal dimensions. Two such dimensions were identified as being required to represent the simple auditory pitch patterns that have been the focus of the present research (monophonic, atonal melodies) – a scalar dimension, which represents relative pitch, and a temporal dimension, which represents the relative timing of auditory events. This supramodal pattern space was labelled a 1½-D space to reflect the qualitative distinction between the dimensions (in terms of their directionality) from which it is constructed.

The SPS framework provided a means of comparing the processing of equivalent (or non-equivalent) structural information presented in different sensory modalities (auditory and visual), and was tested by examining the perception of pattern regularities described by two special types of isomorphic structural transformation: inverse and retrograde. For patterns represented in a 1½-D supramodal pattern space, the perception of inverse transformations requires an inversion of ordinal relations on the scalar dimension, whilst the perception of retrograde transformations requires an inversion of ordinal relations on the temporal dimension. One of the assumptions of the SPS framework was that inversions on the temporal dimension would be harder to process, due to the dimension's inherent directionality. Thus, the main hypothesis examined in all experiments reported in Chapters 4, 5 and 6 was that when stimuli corresponded to representations in a 1½-D supramodal pattern space, structural regularities would be perceived more effectively when they are described by inverse compared to retrograde transformation, irrespective of the sensory modality from which structural information has been abstracted (see Chapter 2, Table 2.3 for a summary of the 1½-D hypothesis and its assumptions).

Before the 1½-D hypothesis was tested, an experiment was first carried out that investigated the possibility that auditory pitch space and visual space share a common metric (Experiment 1, Chapter 3). This possibility has received little (or no) attention in previous experiments that have also made a structural analogy between auditory pitch and visuo-spatial patterns. The experiment used a cross-modal scaling paradigm (adapted from an earlier study carried out by Mudd, 1963). In the experimental trials, participants were presented with a reference tone that was represented by an accompanying reference object, positioned at the centre of a computer screen. The reference tone and object were followed by a comparison tone of varying pitch (taken from a 5-note equal temperament scale). The task was to place a new comparison object anywhere on the screen to represent the comparison tone.

Analysis of the results revealed a general tendency for higher-pitched comparison tones to be represented above and to the right of the reference object, and lower-pitched comparison tones to be represented below and to the left of the reference object (though there were some individual differences). Furthermore, the distance between the reference and comparison objects increased with increasing pitch distance between the reference and comparison tones. The key finding, however, was that an increase of one interval of the 5-note equal temperament scale predicted an increase in visual distance between reference and comparison objects of 1.35° visual angle. As a result of this finding, all subsequent experiments used analogous auditory and visual stimuli in which pitch interval corresponded to visual angle at the ratio of 1:1.35.

The main body of the research consisted of 7 experiments, which were reported in Chapters 4, 5 and 6 (see Table 7.1 for a summary). The paradigm adopted for the 5 experiments reported in Chapters 4 and 5 was a short-term recognition paradigm, which required participants to identify when target patterns were a transformation (inverse, retrograde) of a preceding standard pattern, and when they were not.

In Chapter 4, transformation recognition was examined in two experiments (Experiment 2 and 3), when auditory and visual stimuli corresponded to non-equivalent supramodal pattern spaces (1½-D versus 2½-D). Previous research that has made a structural analogy between auditory pitch and visuo-spatial patterns has typically used visual stimuli that map the pitch of tones onto the vertical axis and the timing of tones onto the horizontal axis. As a starting point, the visual stimuli used in Chapter 4 were presented in this way. According to the SPS framework, sequential patterns of visual objects presented at different positions on vertical and horizontal axes correspond to a 2½-D supramodal pattern space, constructed from two scalar dimensions and a temporal dimension. For patterns that correspond to a 2½-D supramodal pattern space, inverse transformations require an inversion of ordinal relations on a scalar dimension, whilst retrograde transformations also require an inversion on a scalar dimension and/or a temporal dimension. Therefore, an additional 2½-D hypothesis was made which predicted that inverse transformations would be perceived no more effectively than retrograde transformations, owing to the fact that both types of transformation require an inversion of ordinal relations on an equivalent scalar dimension. Furthermore, it predicted that retrograde transformations might be perceived more effectively, owing to the additional structural redundancy on the temporal dimension.

Experiment 2 presented patterns in unimodal trials. In the auditory condition the 1½-D hypothesis was unsupported, as retrograde transformations were recognised more successfully than inverse transformations. By taking into consideration the way in which target stimuli were presented, it was proposed that the result could have been influenced by the processing of redundant non-structural information, which facilitated the recognition of retrograde transformations. In contrast, in the visual condition there was no effect of transformation, which was in agreement with the prediction made by the 2½-D hypothesis. Experiment 3 presented patterns in cross-modal trials. Once more, contrasting effects of transformation were found in different modality conditions. In the AV condition, inverse transformations were recognised more successfully, which was in-line with the prediction made by the 1½-D hypothesis. In the VA condition retrograde transformations were recognised more successfully, which was in-line with the prediction made by the 2½-D hypothesis. It was concluded that recognition was based on the processing of structural information, abstracted from standard patterns, in anticipation of the target.

In Chapter 5, transformation recognition was examined in three experiments (Experiments 4, 5 and 6) when auditory and visual stimuli corresponded to equivalent supramodal pattern spaces (1½-D). Experiment 4 presented patterns in unimodal trials. The results failed to support the 1½-D hypothesis, as no effects of transformation were observed. Once again, it was proposed that the result could have been influenced by the processing of redundant non-structural information, which facilitated the recognition of retrograde transformations in both modality conditions. Experiment 5 presented patterns in cross-modal trials. The 1½-D hypothesis was supported in the AV condition (replicating the results found in Experiment 3), as inverse transformations were recognised more successfully than retrograde transformations. However, the 1½-D hypothesis was not supported by the results in the VA condition, which failed to reveal an effect of transformation. It was argued that this may have been due to participants adopting 'visualisation' strategies, which facilitated the processing of retrograde transformations.

Experiment 6 presented patterns in hybrid trials: auditory or visual standards were followed by bimodal targets. In order to address the possible contaminating effects of non-structural information and visualisation strategies, and to ensure that recognition could only be based on the processing of structural information, all targets were transposed. The key finding was that the 1½-D hypothesis received some support in both modality conditions; less error was made when recognising inverse transformations, whether the standard was auditory or visual. However, this support was not definitive, as the analysis on RT and d' data failed to reveal a transformation effect in the AS condition.

The experiments reported in Chapter 6 (Experiments 7 and 8) also investigated the perception of structural transformations of auditory and visual stimuli that corresponded to representations in 1½-D space. However, a different paradigm was employed to see if transformations would be processed when participants had neither been explicitly instructed to compare patterns, nor had been informed of the way in which the patterns they encountered might be related. The main transformation hypothesis predicted that relationships between patterns under inverse and retrograde structural transformations would be perceived in unimodal and cross-modal conditions. The 1½-D hypothesis, which predicted that inverse transformations would be perceived more effectively than retrograde transformations, was also tested.

A structural priming paradigm was used that comprised experimental trials in which target patterns were preceded by structurally related (inverse, retrograde) or unrelated prime patterns. Participants completed a simple perceptual task that required them to compare the final two elements of the target pattern: for auditory target patterns this was a pitch comparison task; for visual targets this was a spatial comparison task. Any facilitation effects of related primes were interpreted as evidence for the perception of structural relationships between the prime and target patterns. The inter-stimulus interval (ISI) between patterns was also varied to explore the time-courses of hypothesised supramodal mechanisms.

Experiment 7 presented auditory patterns in unimodal trials. Responses were facilitated when patterns were related under retrograde transformation, providing some support for the transformation hypothesis. However, no facilitation effects were observed when patterns were related under inverse transformation. There were no effects of ISI, which indicated that expectations were relatively enduring. Experiment 8 presented patterns in cross-modal trials, with visual targets preceded by auditory primes. The pattern of results was the same as in Experiment 7 – responses were facilitated by retrograde transformations but not by inverse transformations. A general effect of ISI was also observed, with fewer errors made as ISI increased. This was interpreted as indicating that the auditory prime impaired visual perceptual responses, in parallel with the more specific facilitation effect of retrograde transformation.

Table 7.1

Chapters 4, 5 and 6: Summary of experiments, hypotheses and results

Experiment	Paradigm[a]	Modality of trials[b]		Visual stimuli	Transposed target	Hypothesis tested	Result
2	STR	Unimodal	A	2½-D	No	1½-D	Not supported
			V			2½-D	Supported
3	STR	Cross-modal	AV	2½-D	n/a	1½-D and 2½-D	Some support for 1½-D hypothesis
			VA			1½-D and 2½-D	2½-D hypothesis supported
4	STR	Unimodal	A	1½-D	No	1½-D	Not supported
			V			1½-D	Not supported
5	STR	Cross-modal	AV	1½-D	n/a	1½-D	Some support
			VA			1½-D	Not supported
6	STR	Hybrid	AS	1½-D	Yes	1½-D	Some support
			VS			1½-D	Supported
7	SP	Unimodal	A	n/a	Yes	Transformation priming	Some support
						1½-D	Not supported
8	SP	Cross-modal	AV	1½-D	n/a	Transformation priming	Some support
						1½-D	Not supported

Note. [a] STR = short-term recognition, SP = structural priming; [b] A = auditory; V = visual; AV = auditory-visual; VA = visual-auditory; AS = auditory standard; VS = visual standard

1.1.1 Assessment of the SPS framework

In this section, the results of the experiments reported in Chapters 4, 5 and 6 (summarised in Table 7.1) will be considered together (where possible), in relation to the main 1½-D hypothesis of the SPS framework.

Support for the main 1½-D hypothesis, which predicted a processing advantage for inverse transformations of structural information, was mixed. In experiments adopting the short-term recognition paradigm, the predicted processing advantage for inverse transformations was observed in four of the eight experimental conditions in which the hypothesis was tested (Experiment 3, AV condition; Experiment 5, AV condition; Experiment 6, AS and VS conditions). In the other four conditions there was either no effect of transformation (Experiment 4, auditory and visual conditions), or a processing advantage for retrograde transformations (Experiment 2, auditory condition; Experiment 3, VA condition). Two issues were highlighted that may have influenced the results in these conditions. Firstly, in unimodal experiments the recognition of retrograde transformations may have been facilitated by the processing of redundant non-structural information (Experiment 2, auditory condition; Experiment 4, auditory and visual condition). Secondly, in cross-modal experiments 'visualisation' strategies may have been used that also facilitated the recognition of retrograde transformations (Experiment 3, VA condition; Experiment 5, VA condition).

An explanation based on the influence of the issues outlined above is unavoidably speculative, and requires further investigation. However, it had some confirmation when the results of Experiment 6, which controlled for the influence of non-structural information and visualisation strategies, revealed a processing advantage for inverse transformations in both modality conditions. In summary, when taken together, it is argued that a processing advantage for inverse transformations was observed when recognition was based only on the processing of structural information, but the advantage was cancelled out when the recognition of retrograde transformations was facilitated by the processing of redundant non-structural information or by the use of alternative 'visualisation' strategies. Following this argument, it may be concluded that the assumptions of the 1½-D hypothesis received some support. However, further experiments are required to confirm the processing advantage for inverse transformations of structural information, represented in 1½-D supramodal pattern spaces.

In contrast, the experiments reported in Chapter 6, which used the structural priming paradigm, failed to support the 1½-D hypothesis as they only provided evidence for perception of retrograde transformations (Experiments 7 and 8). The processing advantage for retrograde transformations could not be explained by the availability of non-structural information, or 'visualisation' strategies, as both issues had been controlled for. Taken in isolation these results could mean that inversions on the scalar dimension are actually processed less effectively than inversions on the temporal dimension – though it is not immediately clear why this should be the case. Taken in consideration with the results of the recognition experiments, they are harder to interpret. If a processing advantage for inverse transformations can be observed in recognition experiments, why would they not be perceived in priming experiments, when retrograde transformations are?

The contrasting pattern of results suggests paradigm specificity (Meiser, 2011); i.e. different cognitive strategies may have been used depending on the specific task requirements. In recognition experiments the task was to make same/different judgements when comparing standard and target patterns, and to perform this task adequately (when non-structural information and 'visualisation' strategies could not be used) a mental transformation of the pattern structure was required. Thus, experiments using this paradigm probably engaged conscious processing mechanisms that actively transformed structural information (the operations of a hypothesised mechanism have been proposed in Chapter 2, see Figure 2.3). In priming experiments, on the other hand, the task was to make comparison judgements between the final two elements of target patterns. This task could be performed without having to mentally transform pattern structure. Had the results revealed no facilitation effects of structural transformations, then it might have been concluded that the paradigm did not engage any processing mechanisms that transform structural information. As the results demonstrate that retrograde transformations were perceived, another conclusion might be that the paradigm engaged a processing mechanism that actively transformed structural information on the temporal dimension, but not on the scalar dimension. Whereas the perception of transformations in short-term recognition experiments involved conscious processes, the perception of retrograde transformations in priming experiments may have involved a combination of conscious and non-conscious processes.

It is unlikely that the perceptual system is unequipped with the means to process patterns that are related under inverse transformation, in the absence of instruction and training (as was provided in recognition experiments). Inverse, along with retrograde, transformations belong to a limited set of regularities that, in isolation or in combination with others, can describe the structure of all manner of possible sequential patterns that are encountered and perceived by an organism (Jones, 1974, 1978). As discussed in Chapter 1 of this thesis (see Section 1.4), it is generally understood that the perceptual system is driven towards the detection of structural regularities in the environment. One of the reasons for this is that structural regularities signal redundancy, which can be discarded from representations and from the need to process repeatedly, and this in turn saves processing cost (Aksentijevic & Gibson, 2012b). By not processing the structural regularities described by inverse transformation, a vast amount of structural redundancy would be unusable. It is more likely that the mechanisms responsible for the processing of inverse transformations were either not engaged by the experimental paradigm, or the effects of perception were not revealed by performance in the specific experimental task.

In conclusion, it is clear that the assumptions of the 1½-D hypothesis were moderately successful in explaining the pattern of results when the perception of structural transformations was based on conscious processes in recognition experiments. However, they were less successful in explaining the pattern of results in priming experiments. This suggests that different cognitive strategies were used depending on the task requirements. A revision of the SPS framework is required to account for the processing of supramodal structural information in different contexts.

1.2 Methodological and conceptual limitations

It is important to consider the findings reported in this thesis in the context of its methodological strengths and its limitations. One of its strengths was the careful attention it gave to analogous auditory and visual stimuli. For example, the vertical distances between different objects in visual stimuli were determined by participants' mappings of tones onto visual space in Experiment 1. This is not something that has been considered in previous research, and it may be important to consider in future work. It is likely that visual representations of pitch space were influenced both by the scale that was used (5-note equal temperament), and by the dimensions of the computer screen onto which tones were mapped. Therefore, the experiment would have to be repeated to obtain pitch-space to visual-space ratios when different scales and experimental apparatus are used.

The short-term recognition paradigm that was used in Experiments 2 to 6 allowed the systematic investigation of conscious processes involved in the perception of structural transformations in different modality conditions. A general weakness of the short-term recognition paradigm was that it required the experimenter to instruct participants on the nature of the structural transformations which they had to recognise. Although great care was taken to give the same instructions to each participant, and to avoid biasing the way they approached the task, there will inevitably have been an influence of the pre-experimental training on performance.

The structural priming paradigm was purposefully employed to account for some of the weaknesses of the short-term recognition paradigm. One of its strengths was that it did not require the experimenter to give any instructions regarding the types of transformation that were under investigation, and therefore avoided the issues of bias mentioned above. However, a potential weakness was that, in attempting to examine transformation perception indirectly via performance in pitch comparison and spatial comparison tasks, the paradigm was not sufficiently sensitive to the effects of structural transformations. Another weakness concerned the relatively high error rate in Experiment 7 (14.74%), when the task was to compare the pitches of tones. RT was the primary dependent variable in these experiments, but some researchers argue that RT data becomes harder to interpret when the error rate exceeds about 5% (Luce, 1986).

In addition to the methodological limitations mentioned above, it should also be noted that there was a conceptual limitation regarding the distinction made between structural and non-structural information (see Chapter 2, Section 2.2). When applied to auditory stimuli, the term non-structural information has been used to refer to the specific pitch of tones. Tones that shared the same pitch shared the same spectro-temporal properties, and it is known that specific frequencies excite specific receptor cells along the basilar membrane of the cochlea. This frequency-to-place relationship is termed tonotopic organisation and is preserved in the auditory system up to and perhaps beyond the primary auditory cortex (Kaas & Hackett, 2000; Talavage et al., 2004). Thus, non-structural auditory information, as defined in this thesis, is linked to the absolute pitch of tones that is associated with the excitation of a specific network of neurons in the auditory system.

When applied to visual stimuli, the term non-structural information has been used to refer to the specific positions of objects on the computer screen. If participants had been using a chin-rest, and were instructed to remain focussed on a fixation point as visual stimuli were presented, then the positions of objects on the screen would correspond to specific areas of the visual field, and hence specific sensory receptors on the retina of the eye. Just as tonotopic organisation is preserved as information ascends the auditory pathway, retinotopic organisation is also preserved as information ascends the visual pathway (Grill-Spector &

Malach, 2004). Therefore, if a chin-rest and fixation point had been used, the spatial positions of visual objects on the screen would have corresponded to spatial locations in the visual field, which in turn are associated with the excitation of a specific network of neurons in the visual system. However, participants were free to focus their attention anywhere on the screen as visual stimuli were presented. Thus, non-structural information did not correspond to specific spatial locations in the visual field, but to localisation ‘anchors’ in external space.

In short, the term non-structural information was used to refer to the absolute pitch of tones, which is linked to the excitation of a specific network of neurons. But when applied to visual stimuli, it was used to refer to representations of spatial position that did not correspond to absolute spatial locations in the visual field. In highlighting this conceptual limitation, it also becomes apparent that auditory and visual stimuli were not equivalent at the physical and sensory level. Auditory sensory neurons are excited by mechanical vibration, and visual sensory neurons are excited by electromagnetic energy. Visual stimuli comprised black objects on a white background, but a more strictly analogous visual stimulus would comprise white objects on a black background.

1.3 Broader implications

Classic research in sequential pattern learning has identified a number of structural rules that can be used to represent complex sequences (Collard & Povel, 1982; Deutsch & Feroe, 1981; Jones & Zamostny, 1975; Kotovsky & Simon, 1973; Leeuwenberg, 1969; Restle & Brown, 1970; Restle, 1970, 1976; Vitz & Todd, 1969). Although the theories developed by these researchers are now over

40 years old, there have been few developments in this area and as such they remain influential (Fountain & Rowan, 1995; Kundey & Rowan, 2014; Kundey et al., 2013). Rules that correspond to inverse and retrograde transformations feature prominently in these theories.

The present research confirms that the pattern regularities described by these transformations can be perceived in different modality conditions. However, whilst it has been suggested that pattern rules that correspond to retrograde transformations may be used less readily than rules that correspond to inverse transformation (Restle, 1976), the results of Experiment 7 and 8 suggest otherwise. The finding that retrograde transformations were perceived when inverse transformations were not suggests that, when the task does not demand that participants directly compare patterns, regularities described by inverse transformation may not be detected as readily as those described by retrograde transformation.

Inverse and retrograde transformations have also been studied in the music psychology literature. Experiments that have compared the processing of inverse and retrograde transformations have reported contrasting effects – for example, Dowling (1972) found that inverse transformations were processed more effectively than retrograde transformations, but the opposite was found by Cupchik, Phillips and Hill (2001). The present thesis demonstrates that the reason for this might have been due to the way in which stimuli were presented (either with or without additional transposition), and that when retrograde transformations are processed more successfully than inverse transformations, perception might not reflect purely structural processes. Any future experiments investigating melodic transformations clearly need to consider this issue.

The present research has more general implications for melodic perception. Different psychologists have taken the contrasting views that musical ability is either the product of general-purpose cognitive architecture (Bregman, 1990; Handel, 1993), or the product of specialised, music-specific cognitive modules (Peretz & Coltheart, 2003; Schmithorst, 2005). The view promoted here is that at least some of the fundamental processes underlying the perception of melodic structure might not be specific to music, and may not even be specific to the auditory domain. This perspective is shared by other researchers who have investigated the processing of simple melodies and structurally analogous visuo-spatial patterns (Balch & Muscatelli, 1986; McLachlan, Greco, Toner, & Wilson, 2010; Prince, Schmuckler, & Thompson, 2009). However, as it has been noted elsewhere in this thesis, there has been little consensus on how to present structurally analogous visual stimuli. The SPS framework may prove useful in this context as it provides a means for designing stimuli in different sensory modalities that correspond to equivalent structural representations.

Finally, the notion of supramodal structural processing has important implications for functional organisation in the brain. It is generally accepted that information processing in the brain is hierarchical and functionally specialised. Processing streams have been identified that project from sensory cortical centres towards higher-order amodal cortical areas (Creem & Proffitt, 2001; Rauschecker, 2013). The posterior parietal cortex receives projections from both the visual and the auditory cortex, and is known to be concerned with the integration of multimodal information for constructing a spatial representation of the external world (Grefkes & Fink, 2005). However, this area has recently been associated with the processing of retrograde transformations of melody (Foster, Halpern, &

Zatorre, 2013; Foster & Zatorre, 2010; Zatorre, Halpern, & Bouffard, 2010). In light of the present research, it may be hypothesised that one of the functions of the cortex, and in particular this area of the cortex, is to process supramodal structural information that has been abstracted from sensory information in the primary sensory cortices.

1.4 Future directions

The present research represents only the initiation of a research programme concerned with the potential supramodal representation and processing of structural information. As such, the results have prompted a greater number of questions than they have provided answers. These questions invite further investigation, and some possible future directions will be briefly discussed here. In addition to these suggestions, it would also be important to try and replicate the effects observed in the present research. Too often, strong claims are made on the basis of just a few studies, or even on the findings of an isolated experiment. This problem was highlighted by a recent collaborative project that repeated 100 experiments published in top psychology journals and found that they could reproduce only 36% of original findings (Open Science Collaboration, 2015).

The auditory and visual patterns used in the present research were composed of discrete tones and objects. It would be interesting to see if the observed effects persist when continuous stimuli are used instead. Extending the research in this way would provide a route for generalising the present findings to real-life situations. This could be done quite easily by repeating the recognition experiments, and using different stimuli. For priming experiments, the task would have to be adapted in some way. An alternative task could be to compare the beginning and ending pitch of auditory target patterns, or the beginning and ending spatial position of visual target patterns.

Some interesting effects of music training were revealed by the exploratory analyses that were conducted in the present research. However, the pattern of results across experiments was difficult to interpret, and this may have been due to insufficient control of this variable. More meaningful data might be collected by recruiting participants with more extensive and formalised musical training (e.g. students from a music conservatoire), and comparing their results with those of participants with no formal music training. Following the general notion that structural information, abstracted from auditory stimuli, is represented and processed supramodally, it might be predicted that any processing advantage associated with music training would be consistently observed in auditory and visual conditions.

The effects of complexity on the processing of structural transformations could also be investigated. Complexity was discussed in Chapter 1 (Section 1.4) of this thesis, but it was not investigated in the reported experiments. Stimulus patterns were designed in such a way that they were relatively similar in terms of their complexity, and yet, supplementary analysis of PE and RT data in related conditions shows that there was a significant effect of pattern type in all experiments, except for Experiments 4 and 8 (see Appendix V). It would be interesting to see which aspects of pattern structure have the greatest effect on perception, and whether the same structural features influence performance in different modality and transformation conditions. For example, local and global structural complexity could be manipulated independently to examine their effects on transformation perception. Given the theoretical considerations discussed in Chapter 1, it might be expected that global structural complexity would affect performance more than local structural complexity. However, it has been suggested by Cupchik, Phillips and Hill (2001) that the perception of retrograde transformations (of melody) involves a more global cognitive strategy than the perception of inverse transformations. This would predict that the local structural complexity would have the greater influence on the processing of inverse transformations, and global structural complexity would have the greater influence on the processing of retrograde transformations.

As mentioned above, recent brain-imaging research has demonstrated that the processing of melodic transformation is associated with activity located in the intraparietal sulcus, an area that has been previously associated with the processing of visuo-spatial information (Foster et al., 2013; Foster & Zatorre, 2010; Zatorre et al., 2010). In light of the present research, it would be of value to investigate the anatomical areas associated with the mental transformation of auditory and visual stimuli that correspond to equivalent and non-equivalent supramodal pattern spaces. It might be expected that when auditory and visual stimuli correspond to equivalent supramodal pattern spaces, there would be a greater overlap of activity in the parietal cortex, compared to when they correspond to non-equivalent supramodal pattern spaces.

1.5 Conclusion

The present thesis set out to investigate the possibility that the perception and cognition of auditory pitch patterns involve a general mechanism (or mechanisms) that is responsible for the processing of supramodal structural information. In order to explore this possibility, a theoretical framework was outlined that conceives of a supramodal pattern space (SPS). Two supramodal dimensions were described: a scalar dimension that represents the relative pitch of tones or the relative spatial position of visual objects, and a temporal dimension that represents the relative timing of auditory or visual events. In order to test the theoretical assumptions of the SPS framework a series of experiments was devised to investigate the perception of two types of isomorphic transformation: inverse and retrograde. The experiments used either a short-term recognition paradigm or a structural priming paradigm.

In general, the results of the experiments were in agreement with the notion of shared supramodal representations and processes. The key finding was that, when auditory and visual patterns corresponded to an equivalent supramodal pattern space, an inverse transformation processing advantage was observed in experiments adopting the short-term recognition paradigm. This supported the assumption that inversions of ordinal relations on the scalar dimension would be processed more effectively than inversions of ordinal structure on the temporal dimension. However, experiments adopting the structural priming paradigm revealed contrasting results – whilst performance in the indirect perceptual task was apparently facilitated by the perception of retrograde transformations, no evidence was found for the perception if inverse transformations. It was concluded that this demonstrated that different cognitive strategies are used depending on the task, and that different mechanisms may contribute to the conscious and non-conscious perception of structural regularities. This finding can be used to further develop more sophisticated models of supramodal structural processing.

The present research contributes to the existing psychological literature in a number of ways. It brings together a number of diverse areas of psychological research and offers a theoretical solution for how they might be related. To the best of the author's knowledge, the experiments reported here represent the first time that the perception of transformed auditory pitch patterns and visuo-spatial patterns has been directly compared, and the first time that structural information has been proposed to correspond to a supramodal pattern space. Although the theoretical framework outlined in the thesis was not found to provide a definitive account of supramodal structural processing, the findings of the present research can be used to develop more sophisticated theories of pattern perception.

APPENDICES

1 Appendix I: Supporting materials

1.1 Briefing, consent form and debriefing (Experiments 2, 3, 5 and 6)

University of Roehampton
London

ETHICS COMMITTEE
EXPERIMENT BRIEFING

Title of Research Project: The Recognition of Transformed Auditory and Visual Patterns

Ethics Committee Reference: PSYC 11/026

We are investigating the recognition of auditory and visual patterns, when they have undergone two types of transformation. In this experiment you will take part in a number of trials in which you will be presented with two patterns, one after the other. Visual patterns will be presented on a computer screen and auditory patterns will be presented via a pair of headphones. In each trial the second pattern will either be a transformed version of the first (in which case it is related) or it will be a different pattern (in which case it is unrelated). Your task will be to indicate whether you think the second pattern is related or unrelated to the first pattern. Responses will be made by pressing a button on a response box.

There are four blocks in the experiment. The task is not easy – **your full attention and concentration will be needed in order to complete the trials properly**. Extensive instruction and training will be provided before you begin the experimental trials, and you will be given breaks in between blocks. The experiment will take approximately 60 minutes to complete.

Investigator Contact Details:
Michael Thorpe
Department of Psychology
Whitelands College
Holybourne Avenue, SW15 4JD
michael.thorpe@roehampton.ac.uk
+44 (0)20 8392 3496

Director of Studies Contact Details:
Michael Eysenck
Whitelands College
Holybourne Avenue, SW15 4JD
michael.eysenck@roehampton.ac.uk
+44 (0)20 8392 3510

Head of Department Contact Details:
Diane Bray
Whitelands College
Holybourne Avenue, SW15 4JD
d.bray@roehampton.ac.uk
+44 (0)20 8392 3627

University of Roehampton
London

Participant number:

ETHICS COMMITTEE
PARTICIPANT CONSENT FORM

Title of Research Project: The Recognition of Transformed Auditory and Visual Patterns

Ethics Committee, Reference: PSYC 11/026

Your right to withdraw

You can withdraw your participation from the whole experiment or a part of it at any point without needing to justify your decision. You can also request for your data to be withdrawn at any time after participation in the study. In order to do this, please contact the investigator with your participant number, which you will find on the Debrief Form. Please be aware, however, that data may still be used in a collated form. Finally, if you are a student who is volunteering for course credits as part of an undergraduate module, please be advised that there will be no adverse consequences in relation to assessment for your degree if you decide to withdraw.

All data will be held securely in password protected computer files and locked filing cabinets. No one outside of the research team will have access to your individual data, and anonymity will be protected at all times. Researchers involved in the study will be unaware of any links between your identity and the data collected. Signed consent forms will be kept separately from all other data. Your identity will not be passed on to anyone who is not involved in this study, and will be protected in the publication of any findings.

Please note: if you have a concern about any aspect of your participation or any other queries please raise this with the investigator. However if you would like to contact an independent party please contact the Head of Department (or if the researcher is a student you can also contact the Director of Studies).

Consent Statement:

I agree to take part in this research, and am aware that I am free to withdraw at any point. I understand that the information I provide will be treated in confidence by the investigator and that my identity will be protected in the publication of any findings.

Name ....

Signature ....

Date ....

If you are happy to be contacted about future research that we are carrying out, please provide your email address below (this will be kept confidential and will not be passed on to any third parties):

Email address: ....

University of Roehampton London

Participant number:

ETHICS COMMITTEE
EXPERIMENT DEBRIEFING

Title of Research Project: The Recognition of Transformed Auditory and Visual Patterns

Ethics Committee Reference: PSYC 11/026

I would like to take this opportunity to thank you for your time and valued participation in this experiment. For those interested in the purpose of the experiment they have taken part in, see below for a brief explanation.

The processing of structural information perceived from the different sensory modalities of vision and audition is traditionally treated separately in psychological research. However, increasing evidence points towards shared representations and processes for auditory and visual information.

Recently, it has been demonstrated that auditory pitch is strongly associated with visual space – we appear to map pitch onto the vertical dimension, with higher pitched sounds corresponding to higher elevation, and lower pitched sounds corresponding to lower elevation (Evans & Treisman, 2010; Lidji, Kolinsky, Lochy, & Morais, 2007). This spatial association between pitch and space extends to more complex, patterned information, and performance in a task involving the recognition of transformed melodies has been found to predict performance in a visual mental rotation task (Cupchik, Phillips, & Hill, 2001). Studies such as these have led to the proposal that auditory pitch patterns and visuo-spatial patterns are processed by shared cognitive mechanisms, and this hypothesis has been supported by brain imaging studies that have identified a specific area that is activated by the processing of auditory and visual transformations (Foster & Zatorre, 2010). However, this evidence is far from conclusive, and the present research seeks to conduct a thorough examination of the cognitive preferences and constraints for visual and pitch pattern structure.

If you have any further questions about the experiment now or in the future, then please don't hesitate to contact me.

Kind regards,

Michael Thorpe
Department of Psychology
Whitelands College
Holybourne Avenue, SW15 4JD
michael.thorpe@roehampton.ac.uk
+44 (0)20 8392 3496

1

References:

Cupchik, G. C., Phillips, K., & Hill, D. S. (2001). Shared processes in spatial rotation and musical permutation. *Brain and Cognition*, *46*(3), 373–382. doi:10.1006/brcg.2001.1295

Evans, K. K., & Treisman, A. M. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, *10*(1:6), 1–12.

Foster, N. E. V, & Zatorre, R. J. (2010). A role for the intraparietal sulcus in transforming musical pitch information. *Cerebral Cortex*, *20*(6), 1350–9. doi:10.1093/cercor/bhp199

Lidji, P., Kolinsky, R., Lochy, A., & Morais, J. (2007). Spatial associations for musical stimuli: a piano in the head? *Journal of Experimental Psychology: Human Perception and Performance*, *33*(5), 1189–1207.

Please note: if you have a concern about any aspect of your participation or any other queries please raise this with the investigator. However if you would like to contact an independent party please contact the Head of Department (or if the researcher is a student you can also contact the Director of Studies).

Director of Studies Contact Details:
Michael Eysenck
Whitelands College
Holybourne Avenue, SW15 4JD
michael.eysenck@roehampton.ac.uk
+44 (0)20 8392 3510

Head of Department Contact Details:
Diane Bray
Whitelands College
Holybourne Avenue, SW15 4JD
d.bray@roehampton.ac.uk
+44 (0)20 8392 3627

1.2 Demographic and music background questionnaire

University of Roehampton London

Participant number:

<u>Demographic and Music Background Questionnaire</u>

Please answer the following questions which will be of value to the present research. All personal information that you provide will be strictly confidential. You may leave blank any questions you do not want to answer.

Where appropriate, please circle your answers

1. **Age** ________________
2. **Gender** Male / Female
3. **Are you predominantly left-handed or right-handed?** Left / Right
4. **Do you have any hearing problems, or any other issues that you think may affect your performance in this experiment?**

   Yes / No / Don't want to answer
5. **If English is not your first language (the language you learned from birth and/or speak the best), then please tell us what is below:**

   ____
6. **Have you had any formal or informal music training?** Yes / No

If you answered "Yes" to Question 6, then please complete the questions below:

7. **How many months/years music training have you had?**

   ____
8. **Can you read musical notation fluently?** Yes / No
9. **What is your main instrument, and what others do you play (including voice)?**

   Main: _________________________ Others: _____________________________________
10. **Please describe in more detail what your musical training has involved (e.g. instrument lessons, singing lessons, performance, music theory, qualifications, etc…):**

    ____
11. **Approximately when was the last time you performed to an audience (either formally or informally)?**

    ____

End of questionnaire.

1.3 Instructions: Experiment 1

"In each trial of the experiment you will hear two sounds, a reference tone followed by a comparison tone, on the headphones provided. When you hear the reference tone, a reference object will appear at the centre of the computer screen. This reference object represents the reference tone. When you hear the comparison tone, your task will be to indicate where on the screen it should be represented, considering that the reference tone is already represented by the reference object at the centre of the screen. You can do this by clicking on the reference object with the mouse and dragging it to where you think it should go. This can be anywhere on the screen, and you can move the object in any direction. However, it is important that you try to be consistent. This process will be repeated a number of times – once you have given each response you will need to press the space bar to clear the screen and continue. Please use the chin rest during the experimental trials. The experiment will take approximately 20 minutes to complete, and you will take part in some practice trials before starting."

1.4 Instructions: Experiment 2

INTRODUCTION

"In the present experiment you will be presented with short "melody-like" patterns on the headphones and/or with short sequential patterns of objects on the computer screen, such as these [play example patterns]. You will need to compare different patterns with each other to determine whether they are RELATED under two special types of transformation, or not."

INTRODUCTION TO TRANSFORMATIONS

Instructions to be given before each block (a or b, depending on block):

a. "In this half of the experiment you will need to recognise when the pattern is an inverse transformation. This is when a pattern is turned upside down, and all ups become downs and all downs become ups [play example patterns followed by inverse transformations]."
b. "In this half of the experiment you will need to recognise when the pattern is a retrograde transformation. This is when a pattern is presented backwards, or in reverse order [play examples followed by retrograde transformations]."

INTRODUCTION TO TRIAL TIMELINE

Instructions to be given before each sub-block (select the appropriate terms from parentheses according to the block and sub-block):

"In this block patterns will be presented (on the screen/on the headphones). The second pattern will either be a (inverse/retrograde) transformation of the first, in which case it is RELATED, or it will not be a transformation of the first, in which case it is UNRELATED. You will need to indicate whether you think the second pattern is RELATED to the first pattern, or UNRELATED using the response box. Press (left/right) for RELATED and press (left/right) for UNRELATED.

A warning beep will indicate the start of each trial. Your response should be based on the whole pattern, not just the beginning or the end. Give your response once the second pattern has finished – you will have 5 seconds to do this before the experiment moves on to the next trial. Use the index and middle fingers of your dominant hand to give responses. Note that the second pattern is equally likely to be related or unrelated to the first pattern. Before you start there will be some practice trials – you will receive feedback in the practice trials, but you will not receive any feedback when completing the real trials.

Any questions?"

1.5 Instructions: Experiments 3 and 5

INTRODUCTION

"In the present experiment you will be presented with short "melody-like" patterns on the headphones and/or with short sequential patterns of objects on the computer screen, such as these [play example patterns]. You will need to compare different patterns with each other to determine whether they are RELATED under two special types of transformation, or not."

INTRODUCTION TO TRANSFORMATIONS

Instructions to be given before each block (a or b, depending on block):

a. "In this half of the experiment you will need to recognise when the pattern is an inverse transformation. This is when a pattern is turned upside down, and all ups become downs and all downs become ups [play example patterns followed by inverse transformations]."
b. "In this half of the experiment you will need to recognise when the pattern is a retrograde transformation. This is when a pattern is presented backwards, or in reverse order [play examples followed by retrograde transformations]."

INTRODUCTION TO TRIAL TIMELINE

Instructions to be given before each sub-block (select the appropriate terms from parentheses according to the block and sub-block):

"In this block the first pattern will be presented (on the screen/on the headphones) and the second pattern will be presented (on the screen/on the headphones). The second pattern will either be a (inverse/retrograde) transformation of the first, in which case it is RELATED, or it will not be a transformation of the first, in which case it is UNRELATED. You will need to indicate whether you think the second pattern is RELATED to the first pattern, or UNRELATED using the response box. Press (left/right) for RELATED and press (left/right) for UNRELATED.

A warning beep will indicate the start of each trial. Your response should be based on the whole pattern, not just the beginning or the end. Give your response once the second pattern has finished – you will have 5 seconds to do this before the experiment moves on to the next trial. Use the index and middle fingers of your dominant hand to give responses. Note that the second pattern is equally likely to be related or unrelated to the first pattern. Before you start there will be some practice trials – you will receive feedback in the practice trials, but you will not receive any feedback when completing the real trials.

Any questions?"

1.6 Instructions: Experiment 4

INTRODUCTION *(select appropriate term from parentheses depending on the modality condition of the experimental session)*

"In the present experiment you will be presented with (auditory/visual) patterns such as this [play example pattern]. You will need to compare different patterns with each other to determine whether they are RELATED under two special types of transformation, or not."

"In the present experiment you will be presented with short sequential patterns such as these [play example pattern]. You will need to compare different patterns with each other to determine whether they are RELATED under two special types of transformation, or not."

INTRODUCTION TO TRANSFORMATIONS

Instructions to be given before each block (a or b, depending on block – select appropriate terms from parentheses):

a. "In this half of the experiment you will need to recognise when the pattern is an inverse transformation. This is when a pattern is turned upside down, and all ups become downs and all downs become ups [play example patterns followed by inverse transformations]."
b. "In this half of the experiment you will need to recognise when the pattern is a retrograde transformation. This is when a pattern is presented backwards, or in reverse order [play example followed by retrograde transformation]."

"In each experimental trial, the second pattern will either be a (inverse/retrograde) transformation of the first, in which case it is RELATED, or it will not be a transformation of the first, in which case it is UNRELATED. You will need to indicate whether you think the second pattern is RELATED to the first pattern, or UNRELATED using the response box. Press (left/right) for RELATED and press (left/right) for UNRELATED.

A warning beep will indicate the start of each trial. Your response should be based on the whole pattern, not just the beginning or the end. Give your response once the second pattern has finished – you will have 5 seconds to do this before the experiment moves on to the next trial. Use the index and middle fingers of your dominant hand to give responses. Note that the second pattern is equally likely to be related or unrelated to the first pattern. Before you start there will be some practice trials – you will receive feedback in the practice trials, but you will not receive any feedback when completing the real trials.

Any questions?"

1.7 Instructions: Experiment 6

INTRODUCTION

"In the present experiment you will be presented with patterns such as this [play example pattern]. You will need to compare different patterns with each other to determine whether they are RELATED under two special types of transformation, or not."

INTRODUCTION TO TRANSFORMATIONS

Instructions to be given before each block (a or b, depending on block):

a. "In this half of the experiment you will need to recognise when the pattern is an inverse transformation. This is when a pattern is turned upside down, and all ups become downs and all downs become ups [play example]."
b. "In this half of the experiment you will need to recognise when the pattern is a retrograde transformation. This is when a pattern is presented backwards, or in reverse order [play example]."

INTRODUCTION TO TRIAL TIMELINE

Instructions to be given before each sub-block (select the appropriate terms from parentheses according to the block and sub-block):

"In this block the first pattern will be presented (on the screen/in the headphones) only. The second pattern will either be a (inverse/retrograde) transformation of the first, in which case it is RELATED, or it will not be a transformation of the first, in which case it is UNRELATED [play example]. You will need to indicate whether you think the second pattern is RELATED to the first pattern, or UNRELATED using the response box. Press (left/right) for RELATED and press (left/right) for UNRELATED.

A warning beep will indicate the start of each trial. Your response should be based on the whole pattern, not just the beginning or the end. Give your response once the second pattern has finished – you will have 5 seconds to do this before the experiment moves on to the next trial. Use the index and middle fingers of your dominant hand to give responses. Note that the second pattern is equally likely to be related or unrelated to the first pattern. Before you start there will be some practice trials – you will receive feedback in the practice trials, but you will not receive any feedback when completing the real trials.

Any questions?"

1.8 Instructions: Experiment 7

INTRODUCTION TO PITCH COMPARISONS

"In the present experiment your task is to compare two tones presented one after the other, and indicate whether the second tone is higher or lower in pitch than the first. You will now hear some examples. After each example, indicate (using the response box provided) whether you think the second tone is higher or lower than the first. Press the left button with the index finger of your left hand for a (lower/higher) response, and the right button with the index finger or your right hand for a (lower/higher) response. You will receive feedback after each response." [Begin examples]

INTRODUCTION TO THE EXPERIMENTAL TRIALS

"In each trial of this experiment you will hear two melodies, presented one after the other. All melodies will always be six tones in length. You need only respond to the last two tones of the second melody.

When a trial begins a fixation point will appear on the screen, accompanied by a short beep. Shortly afterwards, you will be presented with the first melody. There will then be a short pause of varying length, followed by the second melody. You will respond to the last two tones of the second melody. Once you have heard it, you must indicate on the response box whether the final tone was higher or lower in pitch than the preceding tone.

Press the left button with the index finger of your left hand for a (lower/higher) response, and the right button with the index finger or your right hand for a (lower/higher) response.

Please respond as quickly as possible whilst maintaining your accuracy. Once the second melody has finished you will have 2 seconds to respond before the experiment moves on automatically to the next trial. If you do not respond in this time you will hear two beeps [play example] and you will be prompted to try and respond more quickly. You will also be notified if you respond incorrectly. In this case you will hear a different noise [play example] accompanied by an error message on-screen.

The experiment is split into two blocks, though you will be given the chance to take a break at regular intervals within these blocks. Before you start, there will be some practice trials.

Any questions?"

1.9 Instructions: Experiment 8

"In each trial of this experiment you will be presented with a pattern of visual objects appearing at different heights on the screen. Each pattern will be six objects in length. Your task will be to wait until the final object is presented, and to indicate whether it is higher or lower than the object that was presented immediately before it. Before each pattern is presented, you will hear a short melody in the headphones – you do not need to respond to this melody.

When a trial begins a fixation point will appear on the screen, accompanied by a short beep. Shortly afterwards, you will hear the melody. There will then be a short pause of varying length, followed by the visual pattern.

Press the left button with the index finger of your left hand for a (lower/higher) response, and the right button with the index finger or your right hand for a (lower/higher) response.

Please respond as quickly as possible whilst maintaining your accuracy. Once the visual pattern has finished you will have 2 seconds to respond before the experiment moves on automatically to the next trial. If you do not respond in this time you will hear two beeps [play example] and you will be prompted to try and respond more quickly. You will also be notified if you respond incorrectly. In this case you will hear a different noise [play example] accompanied by an error message on-screen.

The experiment is split into two blocks, though you will be given the chance to take a break at regular intervals within these blocks. Before you start, there will be some practice trials.

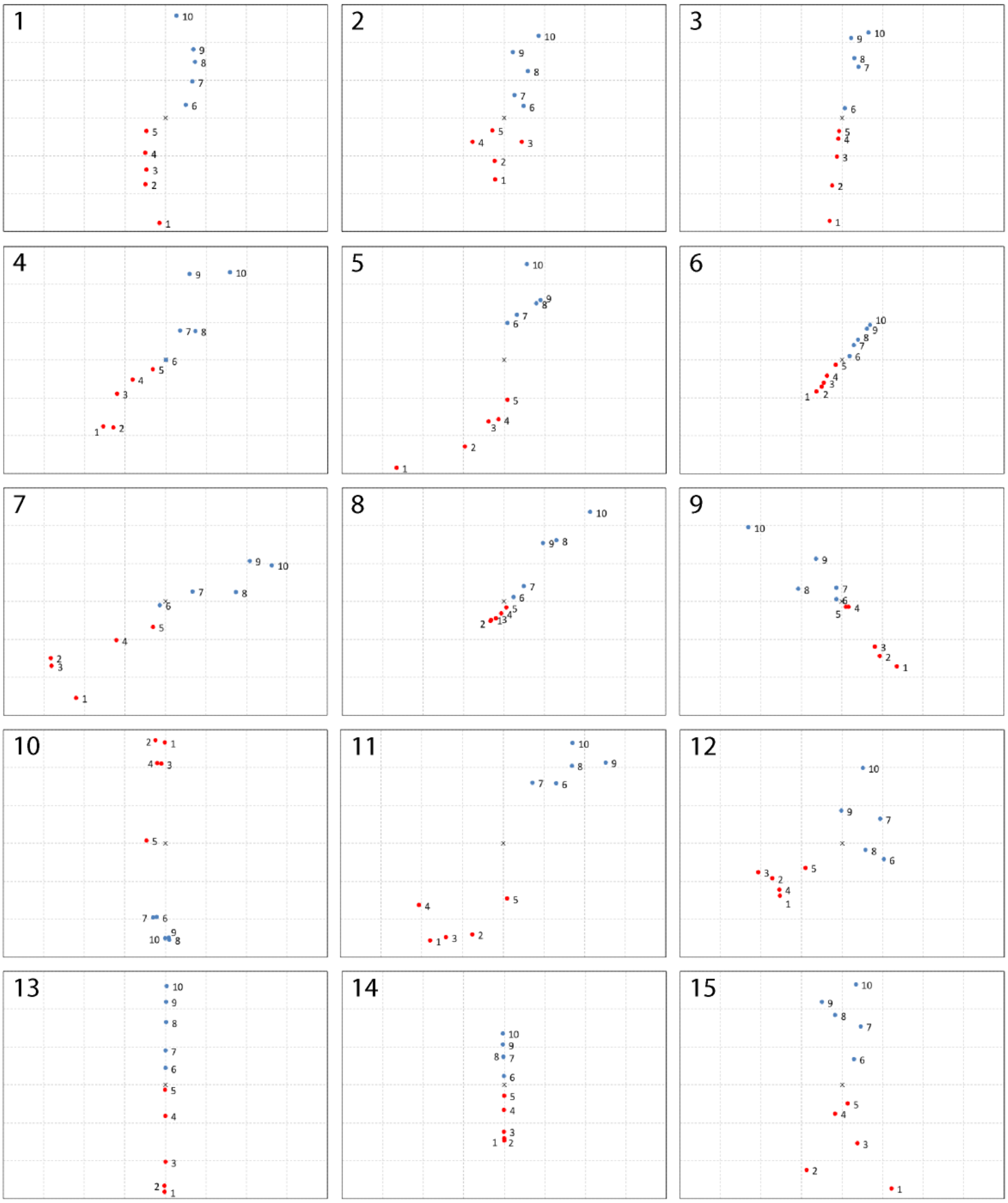
Any questions?"

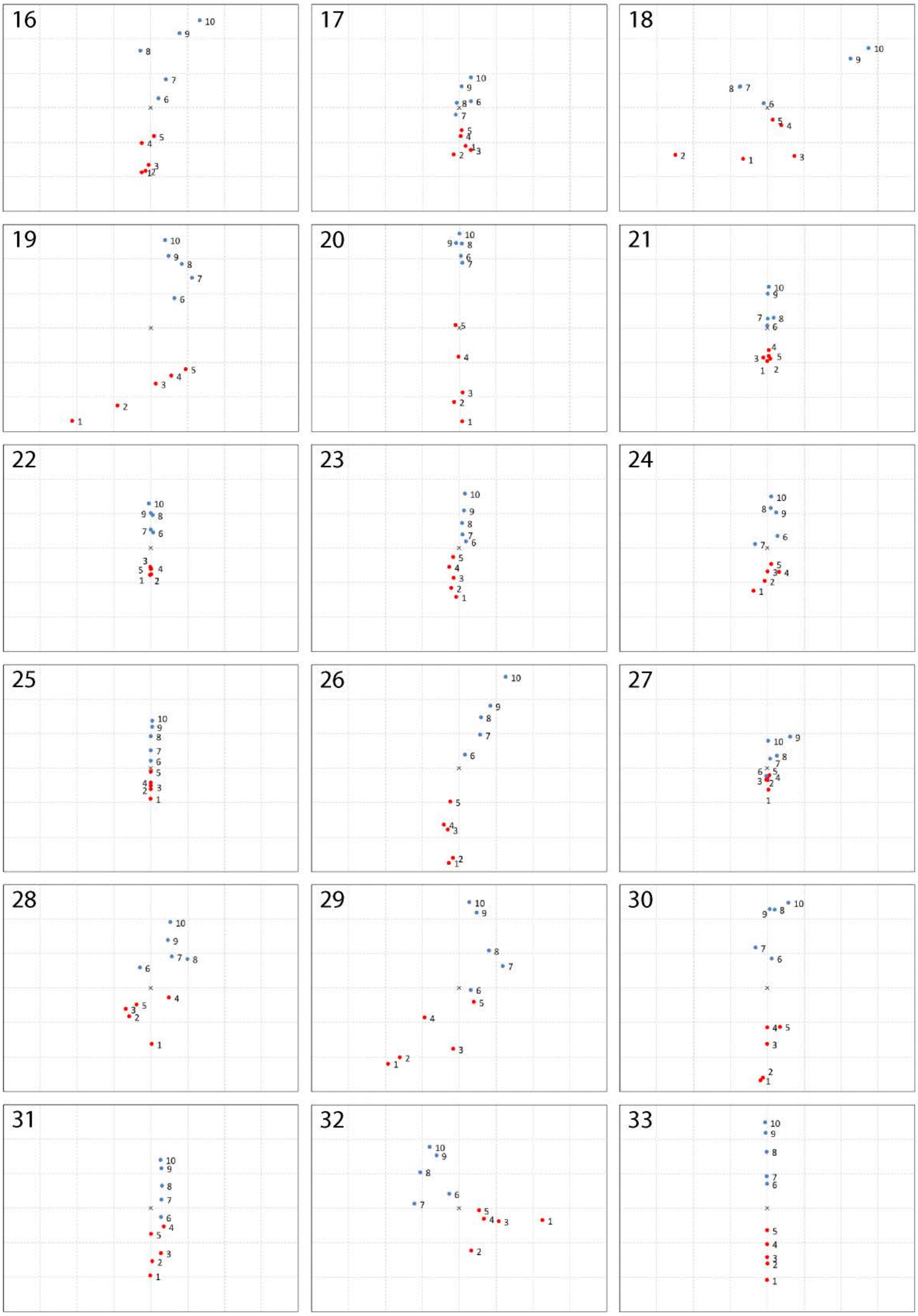
1.10 Ethics statement

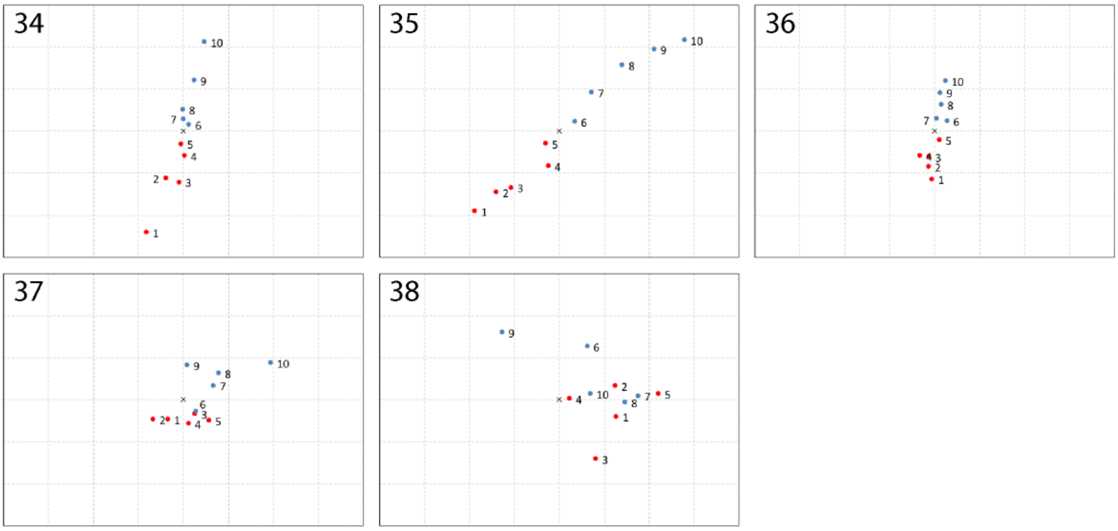
The research for this project was submitted for ethics consideration under the reference PSYC 11/026 in the Department of Psychology and was approved under the procedures of the University of Roehampton's Ethics Committee on 7th September 2011.

2 Appendix II: Individual participant data, Experiment 1

Each figure displays the mean positions of comparison objects for individual participants (total 38). The reference tone (260.00 Hz) was represented by a reference object (x) positioned at the centre of the computer screen (dimensions 25.40 cm x 19.00 cm). Series labels denote comparison tones (red dots: 1 = 130 Hz, 2 = 149.33 Hz, 3 = 171.54 Hz, 4 = 197.04 Hz, 5 = 226.34 Hz; blue dots: 6 = 298.66 Hz, 7 = 343.07 Hz, 8 = 394.09 Hz, 9 = 452.69 Hz, 10 = 520.00 Hz).




16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33


34
35
36
37
38

3 Appendix III: Main analysis

3.1 Experiment 2: Mean data, ANOVA tables and comparisons

PE DATA

Mean PE data

Modality	Transformation	Response	%	*SD*
auditory	inverse	hit	67.67	16.76
		miss	31.82	16.57
		correct rejection	53.33	14.85
		false alarm	45.79	14.68
	retrograde	hit	76.92	15.82
		miss	22.64	15.73
		correct rejection	59.94	17.35
		false alarm	39.12	17.47
visual	inverse	hit	75.09	15.91
		miss	22.52	13.79
		correct rejection	66.29	18.78
		false alarm	31.20	17.33
	retrograde	hit	75.35	17.57
		miss	22.61	15.07
		correct rejection	69.37	19.79
		false alarm	28.24	16.54

Relatedness*Modality

Source	df	MS	F	sig.	η_p^2
relatedness	1	0.84	38.34	0.000	0.42
Error(relatedness)	52	0.02			
modality	1	0.56	33.31	0.000	0.39
Error(modality)	52	0.02			
relatedness*modality	1	0.10	9.76	0.003	0.16
Error(relatedness*modality)	52	0.01			

Relatedness*Modality comparisons

Condition	Comparison	MD	SEM	sig.
auditory	related-unrelated	15.22	2.17	0.000
visual	related-unrelated	7.15	2.11	0.002
related	auditory-visual	4.67	1.92	0.014
unrelated	auditory-visual	12.74	1.92	0.000

Appendix III

Modality*Transformation

Source	df	MS	F	sig.	η_p^2
modality	1	0.16	5.59	0.022	0.10
Error(modality)	52	0.03			
transformation	1	0.24	7.73	0.008	0.13
Error(transformation)	52	0.03			
modality*transformation	1	0.18	6.87	0.011	0.12
Error(modality*transformation)	52	0.03			

Modality*Transformation comparisons

Condition	Comparison	MD	SEM	sig.
auditory	inverse-retrograde	9.18	2.58	0.001
visual	inverse-retrograde	0.09	2.23	0.790
inverse	auditory-visual	9.31	2.64	0.001
retrograde	auditory-visual	0.03	2.43	0.901

RT DATA

Mean RT data

Modality	Transformation	Response	RT (ms)	SD
auditory	inverse	hit	990.81	431.90
		miss	1314.50	691.35
		correct rejection	1065.79	534.91
		false alarm	1195.88	568.55
	retrograde	hit	859.67	361.28
		miss	1187.52	596.84
		correct rejection	889.84	402.95
		false alarm	1171.31	611.84
visual	inverse	hit	786.42	349.81
		miss	1138.58	623.63
		correct rejection	925.54	411.71
		false alarm	1069.80	569.56
	retrograde	hit	720.51	290.68
		miss	1056.24	599.27
		correct rejection	857.84	328.59
		false alarm	1038.37	520.62

Relatedness*Modality

Source	df	MS	F	sig.	η_p^2
relatedness	1	1.03	18.15	0.000	0.26
Error(relatedness)	52	0.06			
modality	1	0.75	13.07	0.001	0.20
Error(modality)	52	0.06			
relatedness*modality	1	0.29	16.53	0.000	0.24
Error(relatedness*modality)	52	0.02			

Relatedness*Modality comparisons

Condition	Comparison	MD	SEM	sig.
auditory	related-unrelated	52.57	34.45	0.075
visual	related-unrelated	138.22	31.22	0.000
related	auditory-visual	171.78	31.33	0.000
unrelated	auditory-visual	86.13	43.10	0.261

Modality*Transformation

Source	df	MS	F	sig.	η_p^2
modality	1	1.96	29.98	0.000	0.37
Error(modality)	52	0.07			
transformation	1	0.46	6.67	0.013	0.11
Error(transformation)	52	0.07			
modality*transformation	1	0.21	2.83	0.098	0.05
Error(modality*transformation)	52	0.08			

Modality*Transformation comparisons

Condition	Comparison	MD	SEM	sig.
auditory	inverse-retrograde	131.14	45.83	0.005
visual	inverse-retrograde	65.90	40.43	0.564

d′ DATA

Mean *d′* data

Modality	Transformation	*d′*	SD
auditory	inverse	0.59	0.67
	retrograde	1.08	0.81
visual	inverse	1.23	0.83
	retrograde	1.34	0.91

Appendix III

Modality*Transformation

Source	df	MS	F	sig.	η_p^2
modality	1	10.75	25.51	0.000	0.33
Error(modality)	52	0.42			
transformation	1	4.70	17.43	0.000	0.25
Error(transformation)	52	0.27			
modality*transformation	1	1.97	7.18	0.010	0.12
Error(modality*transformation)	52	0.28			

Modality*Transformation comparisons

Condition	Comparison	MD	SEM	sig.
auditory	inverse-retrograde	0.49	0.10	0.000
visual	inverse-retrograde	0.11	0.10	0.306
inverse	auditory-visual	0.64	0.11	0.000
retrograde	auditory-visual	0.26	0.12	0.040

c DATA

Mean *c* data

Modality	Transformation	*c*	SD
auditory	inverse	-0.19	0.28
	retrograde	-0.24	0.33
visual	inverse	-0.09	0.31
	retrograde	-0.06	0.31

Modality*Transformation

Source	df	MS	F	sig.	η_p^2
modality	1	1.03	13.99	0.000	0.21
Error(modality)	52	0.07			
transformation	1	0.01	0.14	0.708	0.00
Error(transformation)	52	0.08			
modality*transformation	1	0.09	1.34	0.253	0.03
Error(modality*transformation)	52	0.07			

3.2 Experiment 3: Mean data, ANOVA tables and comparisons

PE DATA

Mean PE data

Modality	Transformation	Response	%	SD
AV	inverse	hit	82.76	12.41
		miss	16.09	12.15
		correct rejection	70.11	16.68
		false alarm	28.05	16.56
	retrograde	hit	75.40	16.91
		miss	23.45	17.33
		correct rejection	69.89	17.78
		false alarm	27.82	18.18
VA	inverse	hit	80.92	15.81
		miss	17.70	14.17
		correct rejection	75.17	23.02
		false alarm	23.68	21.13
	retrograde	hit	85.29	14.30
		miss	14.02	14.40
		correct rejection	77.70	21.48
		false alarm	20.92	21.29

Relatedness*Modality

Source	df	MS	F	sig.	η_p^2
relatedness	1	0.19	5.46	0.027	0.16
Error(relatedness)	28	0.03			
modality	1	0.16	10.68	0.003	0.28
Error(modality)	28	0.02			
relatedness*modality	1	0.00	0.13	0.718	0.01
Error(relatedness*modality)	28	0.02			

Modality*Transformation

Source	df	MS	F	sig.	η_p^2
modality	1	0.13	4.88	0.035	0.15
Error(modality)	28	0.03			
transformation	1	0.00	0.04	0.839	0.00
Error(transformation)	28	0.03			
modality*transformation	1	0.25	13.40	0.001	0.32
Error(modality*transformation)	28	0.02			

Modality*Transformation comparisons

Condition	Comparison	MD	SEM	sig.
AV	inverse-retrograde	7.36	2.55	0.012
VA	inverse-retrograde	3.68	2.72	0.045
inverse	AV-VA	1.61	2.70	0.575
retrograde	AV-VA	9.43	2.42	0.000

RT DATA

Mean RT data

Modality	Transformation	Response	RT (ms)	SD
AV	inverse	hit	1040.26	401.42
		miss	1595.49	851.44
		correct rejection	1238.98	511.44
		false alarm	1526.93	577.05
	retrograde	hit	1192.14	538.72
		miss	1468.96	858.39
		correct rejection	1210.10	488.18
		false alarm	1643.13	637.61
VA	inverse	hit	998.36	424.23
		miss	1686.64	725.70
		correct rejection	1060.03	449.36
		false alarm	1683.19	991.90
	retrograde	hit	859.33	489.34
		miss	1257.98	595.28
		correct rejection	925.61	454.49
		false alarm	1496.19	831.28

Relatedness*Modality

Source	df	MS	F	sig.	η_p^2
relatedness	1	0.40	6.76	0.015	0.20
Error(relatedness)	28	0.06			
modality	1	1.51	20.67	0.000	0.43
Error(modality)	28	0.07			
relatedness*modality	1	0.04	1.02	0.321	0.04
Error(relatedness*modality)	28	0.04			

Modality*Transformation

Source	df	MS	F	sig.	η_p^2
modality	1	1.05	9.83	0.004	0.26
Error(modality)	28	0.11			
transformation	1	0.06	0.70	0.411	0.02
Error(transformation)	28	0.09			
modality*transformation	1	1.15	15.80	0.000	0.36
Error(modality*transformation)	28	0.07			

Modality*Transformation comparisons

Condition	Comparison	MD	SEM	sig.
AV	inverse-retrograde	151.89	68.08	0.077
VA	inverse-retrograde	139.03	47.71	0.001
inverse	AV-VA	41.89	66.88	0.895
retrograde	AV-VA	332.81	92.92	0.000

d′ DATA

Mean *d′* data

Modality	Transformation	*d′*	SD
AV	inverse	1.59	0.82
	retrograde	1.39	0.96
VA	inverse	1.71	1.07
	retrograde	2.06	1.17

Modality*Transformation

Source	df	MS	F	sig.	η_p^2
modality	1	4.56	11.46	0.002	0.29
Error(modality)	28	0.40			
transformation	1	0.17	0.53	0.473	0.02
Error(transformation)	28	0.32			
modality*transformation	1	2.17	12.57	0.001	0.31
Error(modality*transformation)	28	0.17			

Modality*Transformation comparisons

Condition	Comparison	MD	SEM	sig.
AV	inverse-retrograde	0.20	0.11	0.092
VA	inverse-retrograde	0.35	0.15	0.023
inverse	AV-VA	0.12	0.14	0.396
retrograde	AV-VA	0.67	0.14	0.000

c DATA

Mean *c* data

Modality	Transformation	*c*	SD
AV	inverse	-0.17	0.31
	retrograde	-0.04	0.35
VA	inverse	-0.06	0.31
	retrograde	-0.08	0.35

Modality*Transformation

Source	df	MS	F	sig.	η_p^2
modality	1	0.04	0.46	0.505	0.02
Error(modality)	28	0.09			
transformation	1	0.10	1.46	0.238	0.05
Error(transformation)	28	0.07			
modality*transformation	1	0.20	3.04	0.092	0.10
Error(modality*transformation)	28	0.06			

43

3.3 Experiment 4: Mean data, ANOVA tables and comparisons

PE DATA

Mean PE data

Modality	Transformation	Response	%	SD
auditory	inverse	hit	71.60	15.97
		miss	27.41	15.06
		correct rejection	59.51	12.80
		false alarm	39.75	12.87
	retrograde	hit	74.57	11.99
		miss	24.44	12.40
		correct rejection	61.23	16.85
		false alarm	37.78	16.64
visual	inverse	hit	80.69	16.84
		miss	15.40	11.56
		correct rejection	74.02	23.32
		false alarm	21.38	20.27
	retrograde	hit	79.77	16.85
		miss	17.47	13.62
		correct rejection	75.86	18.89
		false alarm	22.53	18.29

Relatedness*Modality

Source	df	MS	F	sig.	η_p^2
relatedness	1	0.35	16.92	0.000	0.24
relatedness*modality	1	0.04	2.02	0.161	0.04
Error(relatedness)	54	0.02			
modality	1	0.76	19.27	0.000	0.26
Error	54	0.04			

Modality*Transformation

Source	df	MS	F	sig.	η_p^2
modality	1	0.55	9.63	0.003	0.15
Error	54	0.06			
transformation	1	0.01	0.22	0.642	0.00
modality*transformation	1	0.01	0.47	0.498	0.01
Error(transformation)	54	0.02			

Modality*Transformation comparisons

Condition	Comparison	MD	SEM	sig.
auditory	inverse-retrograde	2.96	2.93	0.428
visual	inverse-retrograde	2.07	2.82	0.878

RT DATA

Mean RT data

Modality	Transformation	Response	RT (ms)	SD
auditory	inverse	hit	863.99	428.21
		miss	1306.01	684.26
		correct rejection	981.36	548.70
		false alarm	1151.44	540.31
	retrograde	hit	803.09	412.57
		miss	1130.77	616.91
		correct rejection	898.94	577.19
		false alarm	1059.40	452.78
visual	inverse	hit	740.94	379.21
		miss	924.41	617.09
		correct rejection	871.24	552.66
		false alarm	1046.59	747.96
	retrograde	hit	775.34	480.88
		miss	1070.04	580.92
		correct rejection	809.80	460.44
		false alarm	1038.07	562.91

Appendix III

Relatedness*Modality

Source	df	MS	F	sig.	η_p^2
relatedness	1	0.43	13.44	0.001	0.20
relatedness*modality	1	0.02	0.75	0.389	0.01
Error(relatedness)	54	0.03			
modality	1	0.32	0.50	0.481	0.01
Error	54	0.64			

Modality*Transformation

Source	df	MS	F	sig.	η_p^2
modality	1	0.53	0.73	0.397	0.01
Error	54	0.72			
transformation	1	0.00	0.00	0.950	0.00
modality*transformation	1	0.07	0.60	0.441	0.01
Error(transformation)	54	0.12			

Modality*Transformation comparisons

Condition	Comparison	MD	SEM	sig.
auditory	inverse-retrograde	60.90	54.42	0.562
visual	inverse-retrograde	34.40	52.51	0.609

d′ DATA

Mean *d′* data

Modality	Transformation	*d′*	SD
auditory	inverse	0.86	0.68
	retrograde	0.99	0.65
visual	inverse	1.74	0.85
	retrograde	1.74	1.11

Modality*Transformation

Source	df	MS	F	sig.	η_p^2
modality	1	18.46	16.60	0.000	0.24
Error	54	1.11			
transformation	1	0.14	0.41	0.525	0.01
modality*transformation	1	0.14	0.41	0.526	0.01
Error(transformation)	54	0.33			

Modality*Transformation comparisons

Condition	Comparison	MD	SEM	sig.
auditory	inverse-retrograde	0.14	0.16	0.378
visual	inverse-retrograde	0.00	0.15	0.999

c DATA

Mean *c* data

Modality	Transformation	*c*	SD
auditory	inverse	-0.17	0.25
	retrograde	-0.17	0.29
visual	inverse	-0.04	0.43
	retrograde	-0.05	0.29

Modality*Transformation

Source	df	MS	F	sig.	η_p^2
modality	1	0.46	2.98	0.090	0.05
Error	54	0.15			
transformation	1	0.00	0.00	0.958	0.00
modality*transformation	1	0.00	0.02	0.902	0.00
Error(transformation)	54	0.06			

3.4 Experiment 5: Mean data, ANOVA tables and comparisons

PE DATA

Mean PE data

Modality	Transformation	Response	%	SD
AV	inverse	hit	82.63	15.62
		miss	16.32	15.48
		correct rejection	61.75	19.89
		false alarm	36.32	19.65
	retrograde	hit	76.67	15.56
		miss	21.75	15.32
		correct rejection	59.65	21.08
		false alarm	39.12	20.63
VA	inverse	hit	80.53	12.82
		miss	18.77	12.95
		correct rejection	67.19	17.70
		false alarm	31.40	17.87
	retrograde	hit	78.42	16.65
		miss	20.70	16.68
		correct rejection	68.60	21.30
		false alarm	29.65	20.37

Relatedness*Modality

Source	df	MS	F	sig.	η_p^2
relatedness	1	1.08	40.58	0.000	0.52
Error(relatedness)	37	0.03			
modality	1	0.03	2.17	0.150	0.06
Error(modality)	37	0.02			
relatedness*modality	1	0.09	10.15	0.003	0.22
Error(relatedness*modality)	37	0.01			

Relatedness*Modality comparisons

Condition	Comparison	MD	SEM	sig.
AV	related-unrelated	18.68	2.61	0.000
VA	related-unrelated	10.79	2.37	0.000
related	AV-VA	0.70	1.67	0.427
unrelated	AV-VA	7.19	2.09	0.005

Modality*Transformation

Source	df	MS	F	sig.	η_p^2
modality	1	0.01	0.20	0.655	0.01
Error(modality)	37	0.03			
transformation	1	0.09	1.72	0.198	0.04
Error(transformation)	37	0.05			
modality*transformation	1	0.10	2.29	0.138	0.06
Error(modality*transformation)	37	0.04			

Modality*Transformation comparisons

Condition	Comparison	MD	SEM	sig.
AV	inverse-retrograde	5.44	3.37	0.043
VA	inverse-retrograde	1.93	3.24	0.971

RT DATA

Mean RT data

Modality	Transformation	Response	RT (ms)	SD
AV	inverse	hit	889.56	374.32
		miss	1301.65	558.71
		correct rejection	1172.95	497.24
		false alarm	1251.93	547.99
	retrograde	hit	1035.07	396.95
		miss	1428.94	619.39
		correct rejection	1179.13	457.65
		false alarm	1369.38	560.15
VA	inverse	hit	872.91	426.76
		miss	1640.69	1021.26
		correct rejection	1007.77	415.46
		false alarm	1103.13	584.33
	retrograde	hit	815.71	429.51
		miss	1483.43	919.49
		correct rejection	1004.27	459.26
		false alarm	1177.97	763.93

Relatedness*Modality

Source	df	MS	F	sig.	η_p^2
relatedness	1	1.47	35.78	0.000	0.49
Error(relatedness)	37	0.04			
modality	1	1.19	15.69	0.000	0.30
Error(modality)	37	0.08			
relatedness*modality	1	0.01	0.32	0.576	0.01
Error(relatedness*modality)	37	0.03			

Modality*Transformation

Source	df	MS	F	sig.	η_p^2
modality	1	1.00	10.31	0.003	0.22
Error(modality)	37	0.10			
transformation	1	0.09	0.99	0.326	0.03
Error(transformation)	37	0.09			
modality*transformation	1	0.58	10.12	0.003	0.22
Error(modality*transformation)	37	0.06			

Modality*Transformation comparisons

Condition	Comparison	MD	SEM	sig.
AV	inverse-retrograde	145.51	55.96	0.006
VA	inverse-retrograde	57.21	68.33	0.239
inverse	AV-VA	16.64	57.61	0.520
retrograde	AV-VA	219.36	59.60	0.000

d′ DATA

Mean *d′* data

Modality	Transformation	*d′*	SD
AV	inverse	1.40	1.06
	retrograde	1.09	0.99
VA	inverse	1.39	0.80
	retrograde	1.47	1.10

Modality*Transformation

Source	df	MS	F	sig.	η_p^2
modality	1	1.24	3.20	0.082	0.08
Error(modality)	37	0.39			
transformation	1	0.52	0.68	0.414	0.02
Error(transformation)	37	0.77			
modality*transformation	1	1.52	2.94	0.095	0.07
Error(modality*transformation)	37	0.52			

Modality*Transformation comparisons

Condition	Comparison	MD	SEM	sig.
AV	inverse-retrograde	0.32	0.19	0.096
VA	inverse-retrograde	0.08	0.18	0.652

c DATA

Mean *c* data

Modality	Transformation	*c*	SD
AV	inverse	-0.30	0.28
	retrograde	-0.22	0.30
VA	inverse	-0.18	0.29
	retrograde	-0.13	0.31

Modality*Transformation

Source	df	MS	F	sig.	η_p^2
modality	1	0.42	6.38	0.016	0.15
Error(modality)	37	0.07			
transformation	1	0.16	2.64	0.113	0.07
Error(transformation)	37	0.06			
modality*transformation	1	0.01	0.18	0.678	0.01
Error(modality*transformation)	37	0.05			

3.5 Experiment 6: Mean data, ANOVA tables and comparisons

PE DATA

Mean PE data

Modality	Transformation	Response	%	SD
AS	inverse	hit	80.00	17.73
		miss	18.61	17.43
		correct rejection	58.89	17.85
		false alarm	40.56	18.04
	retrograde	hit	72.50	17.13
		miss	27.50	17.13
		correct rejection	62.50	18.57
		false alarm	36.39	18.69
VS	inverse	hit	86.94	11.91
		miss	12.50	11.31
		correct rejection	68.89	20.53
		false alarm	29.72	20.07
	retrograde	hit	73.33	18.67
		miss	25.00	18.44
		correct rejection	65.00	17.97
		false alarm	33.06	17.37

Relatedness*Modality

Source	df	MS	F	sig.	η_p^2
relatedness	1	1.16	46.12	0.000	0.57
Error(relatedness)	35	0.03			
modality	1	0.13	6.93	0.013	0.17
Error(modality)	35	0.02			
relatedness*modality	1	0.00	0.27	0.609	0.01
Error(relatedness*modality)	35	0.02			

Modality*Transformation

Source	df	MS	F	sig.	η_p^2
modality	1	0.10	2.08	0.158	0.06
Error(modality)	35	0.05			
transformation	1	0.93	18.14	0.000	0.34
Error(transformation)	35	0.05			
modality*transformation	1	0.02	1.10	0.302	0.03
Error(modality*transformation)	35	0.02			

Modality*Transformation comparisons

Condition	Comparison	MD	SEM	sig.
AS	inverse-retrograde	8.89	3.42	0.009
VS	inverse-retrograde	12.50	2.86	0.000

RT DATA

Mean RT data

Modality	Transformation	Response	RT (ms)	SD
AS	inverse	hit	1004.65	497.27
		miss	1436.47	830.57
		correct rejection	1017.01	506.95
		false alarm	1411.85	658.60
	retrograde	hit	957.95	408.34
		miss	1384.80	766.94
		correct rejection	868.58	318.65
		false alarm	1311.51	598.82
VS	inverse	hit	926.20	519.43
		miss	1516.98	718.94
		correct rejection	1007.74	494.04
		false alarm	1376.86	661.52
	retrograde	hit	1103.62	539.25
		miss	1252.26	575.44
		correct rejection	948.32	438.84
		false alarm	1392.94	778.73

Relatedness*Modality

Source	df	MS	F	sig.	η_p^2
relatedness	1	0.01	0.19	0.669	0.01
Error(relatedness)	35	0.06			
modality	1	0.01	0.11	0.740	0.00
Error(modality)	35	0.06			
relatedness*modality	1	0.00	0.06	0.810	0.00
Error(relatedness*modality)	35	0.04			

Modality*Transformation

Source	df	MS	F	sig.	η_p^2
modality	1	0.00	0.01	0.909	0.00
Error(modality)	35	0.09			
transformation	1	0.24	1.16	0.288	0.03
Error(transformation)	35	0.21			
modality*transformation	1	0.38	5.56	0.024	0.14
Error(modality*transformation)	35	0.07			

Modality*Transformation comparisons

Condition	Comparison	MD	SEM	sig.
AS	inverse-retrograde	46.70	74.90	0.814
VS	inverse-retrograde	177.42	95.16	0.043
inverse	AS-VS	78.45	64.00	0.115
retrograde	AS-VS	145.67	64.90	0.141

d' DATA

Mean *d'* data

Modality	Transformation	*d'*	SD
AS	inverse	1.13	0.97
	retrograde	0.99	0.88
VS	inverse	1.63	0.86
	retrograde	1.10	0.87

Modality*Transformation

Source	df	MS	F	sig.	η_p^2
modality	1	3.30	8.76	0.006	0.20
Error(modality)	35	0.38			
transformation	1	4.21	9.01	0.005	0.21
Error(transformation)	35	0.47			
modality*transformation	1	1.40	3.81	0.059	0.10
Error(modality*transformation)	35	0.37			

Modality*Transformation comparisons

Condition	Comparison	MD	SEM	sig.
AS	inverse-retrograde	0.15	0.15	0.353
VS	inverse-retrograde	0.54	0.15	0.001

c DATA

Mean *c* data

Modality	Transformation	*c*	SD
AS	inverse	-0.31	0.29
	retrograde	-0.13	0.31
VS	inverse	-0.26	0.31
	retrograde	-0.10	0.31

Modality*Transformation

Source	df	MS	F	sig.	η_p^2
modality	1	0.06	0.83	0.368	0.02
Error(modality)	35	0.07			
transformation	1	1.14	12.51	0.001	0.26
Error(transformation)	35	0.09			
modality*transformation	1	0.00	0.05	0.832	0.00
Error(modality*transformation)	35	0.07			

3.6 Experiment 7: Mean data, ANOVA tables and comparisons

RT DATA

Mean RT data

Target type	ISI	Mean RT (ms)	SD
inverse	500	757.46	158.59
	2000	772.08	212.82
	4000	785.91	202.30
retrograde	500	736.90	172.48
	2000	767.67	203.23
	4000	748.34	209.31
unrelated	500	793.83	212.01
	2000	772.48	165.44
	4000	811.30	217.23

Target type*ISI

Source	df	MS	F	sig.	η_p^2
target type	2	0.01	5.67	0.007	0.21
Error(target type)	42	0.00			
ISI	2	0.01	1.34	0.274	0.06
Error(ISI)	42	0.01			
target type*ISI	4	0.01	0.81	0.522	0.04
Error(target type*ISI)	84	0.01			

Main effect comparisons

Comparison	MD	SEM	sig.
inverse-retrograde	20.84	12.88	0.109
inverse-unrelated	20.72	14.54	0.133
retrograde-unrelated	41.57	9.39	0.001

PE DATA

Mean PE data

Target type	ISI	PE	SD
inverse	500	19.60	11.62
	2000	17.33	13.35
	4000	17.61	13.99
retrograde	500	10.80	12.07
	2000	12.50	9.05
	4000	10.23	9.16
unrelated	500	16.48	13.58
	2000	13.92	10.90
	4000	14.20	13.25

Target type*ISI

Source	df	MS	F	sig.	η_p^2
target type	2	0.05	6.93	0.003	0.25
Error(target type)	42	0.01			
ISI	2	0.01	0.49	0.619	0.02
Error(ISI)	42	0.02			
target type*ISI	4	0.02	1.08	0.373	0.05
Error(target type*ISI)	84	0.02			

Main effect comparisons

Comparison	MD	SEM	sig.
inverse-retrograde	7.01	1.74	0.002
inverse-unrelated	3.31	1.69	0.095
retrograde-unrelated	3.69	1.71	0.057

3.7 Experiment 8: Mean data, ANOVA tables and comparisons

RT DATA

Mean RT data

Target type	ISI	RT (ms)	SD
inverse	500	822.74	208.13
	2000	808.41	239.83
	4000	802.60	210.38
retrograde	500	819.78	211.35
	2000	790.70	191.99
	4000	785.70	213.63
unrelated	500	837.02	221.16
	2000	779.91	223.37
	4000	798.37	191.08

Target type*ISI

Source	df	MS	F	sig.	η_p^2
target type	2	0.00	1.04	0.363	0.05
Error(target type)	44	0.00			
ISI	2	0.03	2.60	0.085	0.11
Error(ISI)	44	0.01			
target type*ISI	4	0.01	1.48	0.217	0.06
Error(target type*ISI)	88	0.01			

Main effect comparisons

Comparison	MD	SEM	sig.
inverse-retrograde	12.52	10.42	0.245
inverse-unrelated	6.15	10.66	0.162
retrograde-unrelated	6.37	10.90	0.856

PE DATA

Mean PE data

Target type	ISI	PE	SD
inverse	500	10.33	9.16
	2000	8.15	9.50
	4000	5.98	6.10
retrograde	500	7.34	9.16
	2000	5.43	8.49
	4000	3.26	6.21
unrelated	500	9.24	8.61
	2000	8.70	7.48
	4000	5.98	8.32

Target type*ISI

Source	df	MS	F	sig.	η_p^2
target type	2	0.05	6.14	0.004	0.22
Error(target type)	44	0.01			
ISI	2	0.15	7.92	0.001	0.27
Error(ISI)	44	0.02			
target type*ISI	4	0.01	0.48	0.747	0.02
Error(target type*ISI)	88	0.02			

Main effect comparisons

Comparison	MD	SEM	sig.
inverse-retrograde	2.81	1.05	0.001
inverse-unrelated	0.18	1.19	0.742
retrograde-unrelated	2.63	1.36	0.021
500-2000[a]	1.54	1.14	0.550
500-4000[a]	3.90	0.91	0.000
2000-4000[a]	2.36	1.08	0.090

Note. [a] A Bonferroni correction was applied

4 Appendix IV: Supplementary analysis 1 – music training

4.1 Correlations with amount of music training: Experiments 2 to 8

Experiment	DV	N	Condition: Auditory[a] / Auditory-Visual[b] / Auditory standard[c]		Condition: Visual[a] / Visual-Auditory[b] / Visual standard[c]	
			Inverse	Retrograde	Inverse	Retrograde
2	PE	20	-.29	-.36	.01	.19
	RT		-.15	.05	.02	.05
	d'		.44	.31	.31	-.17
	c		.05	-.18	.48*	.17
3	PE	18	.07	-.01	-.07	.03
	RT		-.30	-.36	-.35	-.36
	d'		.19	.21	.27	.40
	c		.34	.38	.26	.61**
4	PE	13[d]	.34	-.23	.48	.85**
	RT		.18	.17	.50	.59*
	d'		-.26	.33	-.28	-.62*
	c		.35	-.11	.17	.79**
5	PE	23	-.18	.00	.00	-.46*
	RT		-.09	-.27	-.09	-.40
	d'		.44*	.24	.17	.56**
	c		.43*	.39	.31	.04
6	PE	28	-.41*	-.05	-.29	-.02
	RT		-.37	-.04	-.11	.25
	d'		.46*	.23	.30	-.18
	c		-.19	.25	.05	-.21

Note. [a] Experiments 2 and 4; [b] Experiments 3 and 5; [c] Experiment 6; [d] Number of participants with some previous training in each modality condition (modality was a between-subjects factor)

*p < .05, **p < .01, ***p < .001

Experiment	DV	N	Condition: Inverse	Retrograde	Unrelated
7	RT	6	-.05	-.22	-.13
	PE		-.11	-.50	-.26
8	RT	4	.19	.00	.10
	PE		-.90	-.54	-.84

*p < .05, **p < .01, ***p < .001

4.2 Experiment 2: Mean data, ANOVA tables and comparisons

PE DATA

Mean PE data

Training	Modality	Transformation	PE	SD
no training	auditory	inverse	35.96	17.07
		retrograde	25.76	16.46
	visual	inverse	23.03	13.13
		retrograde	21.77	14.79
some training	auditory	inverse	25.00	13.49
		retrograde	17.50	13.28
	visual	inverse	21.67	15.12
		retrograde	24.00	15.81

Music training*modality*transformation

Source	df	MS	F	sig.	η_p^2
music training	1.00	0.17	1.99	0.164	0.04
Error	51.00	0.09			
music training*modality	1.00	0.15	5.59	0.022	0.10
Error(modality)	51.00	0.03			
music training*transformation	1.00	0.03	0.98	0.327	0.02
Error(transformation)	51.00	0.03			
music training*modality*transformation	1.00	0.00	0.05	0.819	0.00
Error(modality*transformation)	51.00	0.03			

Music training comparisons

Condition	Comparison	MD	SEM	sig.
auditory	no training-some training	9.61	3.52	0.019
visual	no training-some training	0.43	3.42	0.921

Note. A Bonferroni correction was applied to all comparisons

RT DATA

Mean RT data

Training	Modality	Transformation	RT (ms)	SD
no training	auditory	inverse	963.43	408.15
		retrograde	863.80	356.73
	visual	inverse	819.01	341.39
		retrograde	671.67	244.28
some training	auditory	inverse	1036.01	475.93
		retrograde	852.86	377.92
	visual	inverse	732.64	365.69
		retrograde	801.11	346.08

Music training*modality*transformation

Source	df	MS	F	sig.	η_p^2
music training	1.00	0.04	0.07	0.790	0.00
Error	51.00	0.59			
music training*modality	1.00	0.01	0.13	0.725	0.00
Error(modality)	51.00	0.00			
music training*transformation	1.00	0.04	0.54	0.465	0.01
Error(transformation)	51.00	0.00			
music training*modality*transformation	1.00	0.32	4.46	0.040	0.08
Error(modality*transformation)	51.00	0.07			

Music training comparisons

Condition		Comparison	MD	SEM	sig.
auditory	inverse	no training-some training	72.58	123.17	0.933
	retrograde	no training-some training	10.93	103.37	0.462
visual	inverse	no training-some training	86.37	99.36	0.375
	retrograde	no training-some training	129.44	81.18	0.446

Note. A Bonferroni correction was applied to all comparisons

d' DATA

Mean *d'* data

Training	Modality	Transformation	*d'*	SD
no training	auditory	inverse	0.46	0.62
		retrograde	0.92	0.84
	visual	inverse	1.12	0.84
		retrograde	1.37	0.94
some training	auditory	inverse	0.81	0.69
		retrograde	1.34	0.70
	visual	inverse	1.42	0.80
		retrograde	1.29	0.87

Music training*modality*transformation

Source	df	MS	F	sig.	η_p^2
music training	1.00	3.04	1.87	0.177	0.04
Error	51.00	1.62			
music training*modality	1.00	0.97	2.35	0.131	0.04
Error(modality)	51.00	0.41			
music training*transformation	1.00	0.29	1.09	0.301	0.02
Error(transformation)	51.00	0.27			
music training*modality*transformation	1.00	0.60	2.22	0.142	0.04
Error(modality*transformation)	51.00	0.27			

c DATA

Mean *c* data

Training	Modality	Transformation	*c*	SD
no training	auditory	inverse	-0.14	0.30
		retrograde	-0.23	0.33
	visual	inverse	-0.09	0.32
		retrograde	-0.04	0.32
some training	auditory	inverse	-0.26	0.22
		retrograde	-0.26	0.35
	visual	inverse	-0.10	0.30
		retrograde	-0.09	0.28

Music training*modality*transformation

Source	df	MS	F	sig.	η_p^2
music training	1.00	0.15	0.90	0.347	0.02
Error	51.00	0.16			
music training*modality	1.00	0.03	0.35	0.559	0.01
Error(modality)	51.00	0.07			
music training*transformation	1.00	0.01	0.12	0.732	0.00
Error(transformation)	51.00	0.08			
music training*modality*transformation	1.00	0.05	0.79	0.377	0.02
Error(modality*transformation)	51.00	0.07			

4.3 Experiment 3: Mean data, ANOVA tables and comparisons

PE DATA

Mean PE data

Training	Modality	Transformation	PE	SD
no training	AV	inverse	21.21	10.25
		retrograde	28.48	20.46
	VA	inverse	20.00	14.61
		retrograde	18.18	17.41
some training	AV	inverse	12.96	12.41
		retrograde	20.37	14.90
	VA	inverse	16.30	14.14
		retrograde	11.48	12.06

Music training*modality*transformation

Source	df	MS	F	sig.	η_p^2
music training	1.00	0.23	1.83	0.188	0.06
Error(music training)	27.00	0.13			
music training*modality	1.00	0.01	0.30	0.588	0.01
Error(modality)	27.00	0.03			
music training*transformation	1.00	0.00	0.14	0.711	0.01
Error(transformation)	27.00	0.03			
music training*modality*transformation	1.00	0.03	1.55	0.223	0.05
Error(modality*transformation)	27.00	0.02			

RT DATA

Mean RT data

Training	Modality	Transformation	RT (ms)	SD
no training	AV	inverse	1078.11	288.03
		retrograde	1191.40	527.30
	VA	inverse	1112.31	375.18
		retrograde	919.88	531.95
some training	AV	inverse	1017.12	463.80
		retrograde	1192.60	560.76
	VA	inverse	928.73	447.32
		retrograde	822.33	473.41

Music training*modality*transformation

Source	df	MS	F	sig.	η_p^2
music training	1.00	1.38	0.97	0.332	0.04
Error(music training)	27.00	1.41			
music training*modality	1.00	0.03	0.28	0.599	0.01
Error(modality)	27.00	0.00			
music training*transformation	1.00	0.09	1.00	0.325	0.04
Error(transformation)	27.00	0.00			
music training*modality*transformation	1.00	0.02	0.24	0.625	0.01
Error(modality*transformation)	27.00	0.08			

d′ DATA

Mean *d′* data

Training	Modality	Transformation	*d′*	SD
no training	AV	inverse	1.07	0.52
		retrograde	0.98	0.75
	VA	inverse	1.41	1.23
		retrograde	1.67	1.21
some training	AV	inverse	1.90	0.82
		retrograde	1.64	1.01
	VA	inverse	1.89	0.94
		retrograde	2.30	1.10

Music training*modality*transformation

Source	df	MS	F	sig.	η_p^2
music training	1.00	11.52	3.97	0.057	0.13
Error(music training)	27.00	2.91			
music training*modality	1.00	0.25	0.62	0.439	0.02
Error(modality)	27.00	0.40			
music training*transformation	1.00	0.00	0.01	0.936	0.00
Error(transformation)	27.00	0.33			
music training*modality*transformation	1.00	0.19	1.12	0.299	0.04
Error(modality*transformation)	27.00	0.17			

c DATA

Mean *c* data

Training	Modality	Transformation	*c*	SD
no training	AV	inverse	-0.23	0.29
		retrograde	-0.13	0.37
	VA	inverse	-0.08	0.27
		retrograde	-0.17	0.25
some training	AV	inverse	-0.14	0.33
		retrograde	0.02	0.33
	VA	inverse	-0.04	0.34
		retrograde	-0.03	0.39

Music training*modality*transformation

Source	df	MS	F	sig.	η_p^2
music training	1.00	0.28	1.26	0.271	0.05
Error(music training)	27.00	0.22			
music training*modality	1.00	0.01	0.09	0.773	0.00
Error(modality)	27.00	0.09			
music training*transformation	1.00	0.05	0.69	0.414	0.03
Error(transformation)	27.00	0.07			
music training*modality*transformation	1.00	0.01	0.08	0.787	0.00
Error(modality*transformation)	27.00	0.07			

4.4 Experiment 4: Mean data, ANOVA tables and comparisons

PE DATA

Mean PE data

Training	Modality	Transformation	PE	SD
no training	auditory	inverse	25.24	18.15
		retrograde	22.86	14.01
	visual	inverse	14.58	11.98
		retrograde	21.25	14.03
some training	auditory	inverse	29.74	11.09
		retrograde	26.15	10.70
	visual	inverse	16.41	11.42
		retrograde	12.82	12.01

Music training*modality*transformation

Source	df	MS	F	sig.	η_p^2
music training	1.00	0.00	0.00	0.971	0.00
music training*modality	1.00	0.13	2.27	0.138	0.04
Error	52.00	0.06			
music training*transformation	1.00	0.05	2.48	0.121	0.05
music training*modality*transformation	1.00	0.03	1.40	0.242	0.03
Error(transformation)	52.00	0.02			

RT DATA

Mean RT data

Training	Modality	Transformation	RT (ms)	SD
no training	auditory	inverse	756.42	448.79
		retrograde	745.43	456.46
	visual	inverse	825.17	350.97
		retrograde	778.30	471.67
some training	auditory	inverse	979.83	388.67
		retrograde	865.19	367.44
	visual	inverse	637.27	400.58
		retrograde	771.69	511.32

Music training*modality*transformation

Source	df	MS	F	sig.	η_p^2
music training	1.00	0.02	0.03	0.868	0.00
music training*modality	1.00	1.28	1.77	0.189	0.03
Error	52.00	0.72			
music training*transformation	1.00	0.03	0.22	0.644	0.00
music training*modality*transformation	1.00	0.55	4.73	0.034	0.08
Error(transformation)	52.00	0.12			

Music training comparisons

Condition		Comparison	MD	SEM	sig.
auditory	inverse	no training-some training	223.41	152.99	0.170
	retrograde	no training-some training	119.76	175.56	0.595
visual	inverse	no training-some training	187.90	148.32	0.151
	retrograde	no training-some training	6.62	170.19	0.942

Note. A Bonferroni correction was applied to all comparisons

d' DATA

Mean *d'* data

Training	Modality	Transformation	*d'*	SD
no training	auditory	inverse	1.04	0.71
		retrograde	1.07	0.83
	visual	inverse	1.49	0.79
		retrograde	1.46	1.27
some training	auditory	inverse	0.66	0.60
		retrograde	0.91	0.42
	visual	inverse	2.04	0.86
		retrograde	2.08	0.79

Music training*modality*transformation

Source	df	MS	F	sig.	η_p^2
music training	1.00	0.71	0.68	0.413	0.01
music training*modality	1.00	5.07	4.87	0.032	0.09
Error	52.00	1.04			
music training*transformation	0.00	1.00	0.16	0.479	0.49
music training*modality*transformation	0.00	1.00	0.04	0.105	0.75
Error(transformation)	0.00	52.00			

Music training comparisons

Condition	Comparison	MD	SEM	sig.
auditory	no training-some training	0.27	0.28	0.341
visual	no training-some training	0.59	0.27	0.034

Note. A Bonferroni correction was applied to all comparisons

c DATA

Mean *c* data

Training	Modality	Transformation	*c*	SD
no training	auditory	inverse	-0.20	0.29
		retrograde	-0.21	0.24
	visual	inverse	-0.17	0.49
		retrograde	-0.03	0.28
some training	auditory	inverse	-0.14	0.22
		retrograde	-0.12	0.33
	visual	inverse	0.13	0.27
		retrograde	-0.07	0.31

Music training*modality*transformation

Source	df	MS	F	sig.	η_p^2
music training	1.00	0.30	1.94	0.170	0.04
music training*modality	1.00	0.02	0.12	0.733	0.00
Error	52.00	0.15			
music training*transformation	1.00	0.18	3.59	0.064	0.07
music training*modality*transformation	1.00	0.24	5.02	0.029	0.09
Error(transformation)	52.00	0.05			

Music training comparisons

Condition		Comparison	MD	SEM	sig.
auditory	inverse	no training-some training	0.06	0.13	0.631
	retrograde	no training-some training	0.09	0.11	0.414
visual	inverse	no training-some training	0.30	0.13	0.022
	retrograde	no training-some training	0.04	0.11	0.688

Note. A Bonferroni correction was applied to all comparisons

4.5 Experiment 5: Mean data ANOVA tables, and comparisons

PE DATA

Mean PE data

Training	Modality	Transformation	Mean PE	SD
no training	AV	inverse	22.50	17.02
		retrograde	21.67	11.55
	VA	inverse	17.92	13.92
		retrograde	25.42	18.89
some training	AV	inverse	11.82	12.84
		retrograde	21.82	17.84
	VA	inverse	19.39	12.50
		retrograde	17.27	14.35

Music training*modality*transformation

Source	df	MS	F	sig.	η_p^2
music training	1.00	0.14	1.48	0.232	0.04
Error	36.00	0.09			
music training*modality	1.00	0.04	1.45	0.236	0.04
Error(modality)	36.00	0.03			
music training*transformation	1.00	0.01	0.18	0.678	0.01
Error(transformation)	36.00	0.05			
music training*modality*transformation	1.00	0.21	5.55	0.024	0.13
Error(modality*transformation)	36.00	0.04			

Music training comparisons

Condition		Comparison	MD	SEM	sig.
AV	inverse	no training-some training	10.68	4.84	0.062
	retrograde	no training-some training	0.15	5.10	0.639
VA	inverse	no training-some training	1.48	4.31	0.337
	retrograde	no training-some training	8.14	5.39	0.169

Note. A Bonferroni correction was applied to all comparisons

RT DATA

Mean RT data

Training	Modality	Transformation	Mean RT (ms)	SD
no training	AV	inverse	952.45	357.27
		retrograde	1074.61	370.86
	VA	inverse	1006.36	491.43
		retrograde	833.02	512.44
some training	AV	inverse	843.82	387.94
		retrograde	1006.32	421.11
	VA	inverse	775.86	353.47
		retrograde	833.02	512.44

Music training*modality*transformation

Source	df	MS	F	sig.	η_p^2
music training	1.00	1.07	1.58	0.217	0.04
Error	36.00	0.68			
music training*modality	1.00	0.00	0.03	0.859	0.00
Error(modality)	36.00	0.00			
music training*transformation	1.00	0.10	1.17	0.286	0.03
Error(transformation)	36.00	0.00			
music training*modality*transformation	1.00	0.01	0.21	0.652	0.01
Error(modality*transformation)	36.00	0.06			

d′ DATA

Mean *d′* data

Training	Modality	Transformation	Mean *d′*	SD
no training	AV	inverse	0.79	0.87
		retrograde	0.79	0.59
	VA	inverse	1.17	0.74
		retrograde	1.11	1.13
some training	AV	inverse	1.85	0.98
		retrograde	1.30	1.17
	VA	inverse	1.54	0.82
		retrograde	1.73	1.03

Music training*modality*transformation

Source	df	MS	F	sig.	η_p^2
music training	1.00	15.07	7.75	0.008	0.18
Error	36.00	1.94			
music training*modality	1.00	0.83	2.21	0.146	0.06
Error(modality)	36.00	0.38			
music training*transformation	1.00	0.22	0.28	0.603	0.01
Error(transformation)	36.00	0.78			
music training*modality*transformation	1.00	1.48	3.03	0.091	0.08
Error(modality*transformation)	36.00	0.49			

c DATA

Mean *c* data

Training	Modality	Transformation	Mean *c*	SD
no training	AV	inverse	-0.38	0.22
		retrograde	-0.33	0.23
	VA	inverse	-0.33	0.27
		retrograde	-0.17	0.25
some training	AV	inverse	-0.24	0.31
		retrograde	-0.14	0.32
	VA	inverse	-0.07	0.26
		retrograde	-0.10	0.36

Music training*modality*transformation

Source	df	MS	F	sig.	η_p^2
music training	1.00	1.01	6.80	0.013	0.16
Error	36.00	0.15			
music training*modality	1.00	0.00	0.00	0.958	0.00
Error(modality)	36.00	0.07			
music training*transformation	1.00	0.04	0.70	0.410	0.02
Error(transformation)	36.00	0.06			
music training*modality*transformation	1.00	0.13	2.65	0.112	0.07
Error(modality*transformation)	36.00	0.05			

4.6 Experiment 6: Mean data, ANOVA tables and comparisons

PE DATA

Mean PE data

Training	Modality	Transformation	PE	SD
no training	AS	inverse	28.89	14.95
		retrograde	35.56	18.78
	VS	inverse	26.67	13.23
		retrograde	22.22	13.02
some training	AS	inverse	19.81	12.59
		retrograde	11.11	9.23
	VS	inverse	17.41	12.89
		retrograde	26.11	15.40

Music training*modality*transformation

Source	df	MS	F	sig.	η_p^2
music training	1.00	0.70	5.82	0.021	0.15
Error	34.00	0.12			
music training*modality	1.00	0.06	1.43	0.240	0.04
Error(modality)	34.00	0.00			
music training*transformation	1.00	0.36	8.55	0.006	0.20
Error(transformation)	34.00	0.00			
music training*modality*transformation	1.00	0.00	0.21	0.653	0.01
Error(modality*transformation)	34.00	0.02			

Music training comparisons

Condition	Comparison	MD	SEM	sig.
inverse	no training-some training	17.78	4.18	0.001
retrograde	no training-some training	0.56	5.74	0.586

Note. A Bonferroni correction was applied to all comparisons

RT DATA

Mean RT data

Training	Modality	Transformation	RT (ms)	SD
no training	AS	inverse	1067.37	552.48
		retrograde	1038.87	488.78
	VS	inverse	948.18	349.16
		retrograde	1058.32	465.05
some training	AS	inverse	987.92	417.49
		retrograde	931.44	446.89
	VS	inverse	1001.83	444.65
		retrograde	1058.32	465.05

Appendix IV

Music training*modality*transformation

Source	df	MS	F	sig.	η_p^2
music training	1.00	0.05	0.06	0.810	0.00
Error	34.00	0.85			
music training*modality	1.00	0.05	0.56	0.461	0.02
Error(modality)	34.00	0.00			
music training*transformation	1.00	0.51	2.54	0.121	0.07
Error(transformation)	34.00	0.00			
music training*modality*transformation	1.00	0.00	0.00	0.966	0.00
Error(modality*transformation)	34.00	0.07			

d' DATA

Mean *d'* data

Training	Modality	Transformation	*d'*	SD
no training	AS	inverse	0.30	0.60
		retrograde	0.74	0.64
	VS	inverse	1.27	1.00
		retrograde	0.97	0.48
some training	AS	inverse	1.41	0.92
		retrograde	1.07	0.95
	VS	inverse	1.76	0.79
		retrograde	1.14	0.97

Music training*modality*transformation

Source	df	MS	F	sig.	η_p^2
music training	1.00	7.44	4.03	0.053	0.11
Error	34.00	1.85			
music training*modality	1.00	1.01	2.83	0.102	0.08
Error(modality)	34.00	0.36			
music training*transformation	1.00	2.04	4.85	0.035	0.13
Error(transformation)	34.00	0.42			
music training*modality*transformation	1.00	0.37	1.00	0.324	0.03
Error(modality*transformation)	34.00	0.37			

Music training comparisons

Condition	Comparison	MD	SEM	sig.
inverse	no training-some training	0.80	0.30	0.010
retrograde	no training-some training	0.25	0.28	0.385

Note. A Bonferroni correction was applied to all comparisons

c DATA

Mean *c* data

Training	Modality	Transformation	*c*	SD
no training	AS	inverse	-0.15	0.18
		retrograde	-0.13	0.22
	VS	inverse	-0.11	0.19
		retrograde	-0.21	0.31
some training	AS	inverse	-0.37	0.30
		retrograde	-0.13	0.34
	VS	inverse	-0.32	0.33
		retrograde	-0.06	0.31

Music training*modality*transformation

Source	df	MS	F	sig.	η_p^2
music training	1.00	0.12	0.84	0.366	0.02
Error	34.00	0.15			
music training*modality	1.00	0.05	0.65	0.427	0.02
Error(modality)	34.00	0.07			
music training*transformation	1.00	0.56	7.21	0.011	0.18
Error(transformation)	34.00	0.08			
music training*modality*transformation	1.00	0.03	0.45	0.506	0.01
Error(modality*transformation)	34.00	0.07			

Music training comparisons

Condition	Comparison	MD	SEM	sig.
inverse	no training-some training	0.21	0.10	0.033
retrograde	no training-some training	0.08	0.09	0.384

Note. A Bonferroni correction was applied to all comparisons

4.7 Experiment 7: Mean data and ANOVA tables

RT DATA

Mean RT data

Training	Target type	RT (ms)	SD
no training	inverse	813.60	182.46
	retrograde	788.09	193.95
	unrelated	828.79	193.72
some training	inverse	660.39	113.24
	retrograde	651.99	122.59
	unrelated	695.85	141.17

Music training*modality*transformation

Source	df	MS	F	sig.	η_p^2
music training	1	0.51	3.59	0.073	0.15
Error	20	0.14			
music training*transformation	2	0.00	0.07	0.936	0.00
Error(transformation)	40	0.00			

PE DATA

Mean PE data

Training	Target type	PE	SD
no training	inverse	18.88	11.70
	retrograde	12.76	8.67
	unrelated	15.89	11.18
some training	inverse	16.32	10.07
	retrograde	6.94	4.69
	unrelated	12.15	7.62

Music training*modality*transformation

Source	df	MS	F	sig.	η_p^2
music training	1	0.07	0.98	0.333	0.05
Error	20	0.07			
music training*transformation	2	0.00	0.64	0.534	0.03
Error(transformation)	40	0.01			

4.8 Experiment 8: Mean data and ANOVA tables

RT DATA

Mean RT data

Training	Target type	RT (ms)	SD
no training	inverse	835.93	211.17
	retrograde	830.77	190.11
	unrelated	836.72	197.62
some training	inverse	694.00	206.54
	retrograde	646.53	179.35
	unrelated	654.90	204.66

Music training*modality*transformation

Source	df	MS	F	sig.	η_p^2
music training	1	0.67	2.99	0.098	0.13
Error	21	0.22			
music training*transformation	2	0.00	0.77	0.469	0.04
Error(transformation)	42	0.00			

PE DATA

Mean PE data

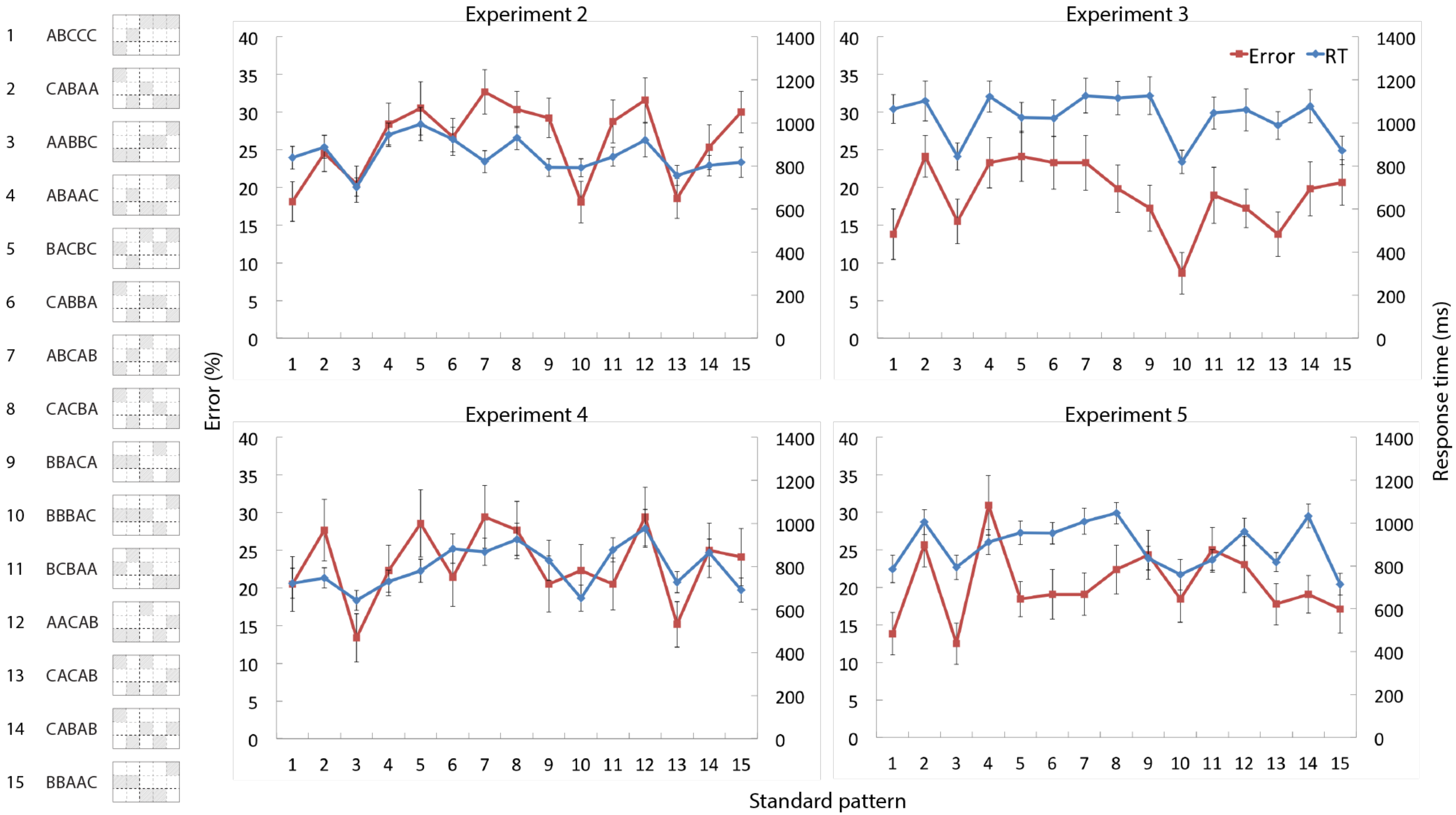
Training	Target type	PE	SD
no training	inverse	6.80	5.51
	retrograde	4.17	5.20
	unrelated	5.81	4.64
some training	inverse	14.58	8.16
	retrograde	10.94	10.40
	unrelated	18.23	7.86

Music training*modality*transformation

Source	df	MS	F	sig.	η_p^2
music training	1	0.25	7.34	0.013	0.26
Error	21	0.03			
music training*transformation	2	0.01	0.95	0.394	0.04
Error(transformation)	42	0.01			

5 Appendix V: Supplementary analysis 2 – complexity

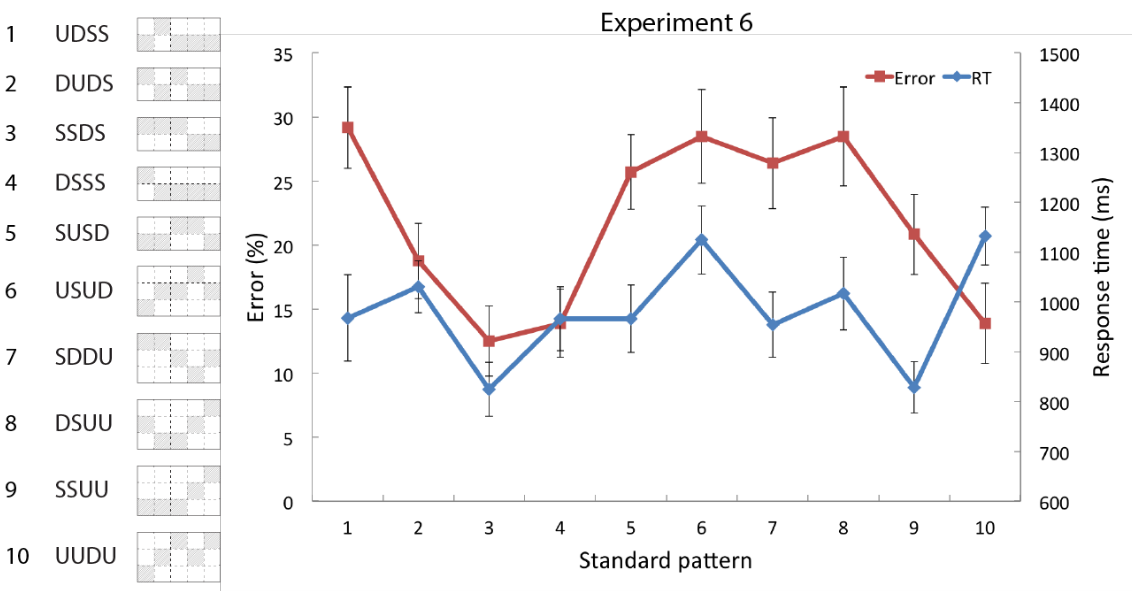
5.1 Experiments 2 to 5: Mean PE and RT in related conditions plotted as a function of standard pattern



ANOVA table – the effects of standard pattern on responses in related trials

Experiment	DV	Source	df	*MS*	*F*	sig.	η_p^2
2	PE	pattern	14.00	0.36	3.65	0.000	0.07
		Error(pattern)	728.00	0.10			
	RT	pattern	9.70	0.79	2.76	0.003	0.06
		Error(pattern)	445.96	0.29			
3	PE	pattern	14.00	0.20	2.44	0.003	0.08
		Error(pattern)	392.00	0.08			
	RT	pattern	14.00	0.41	2.53	0.002	0.09
		Error(pattern)	378.00	0.16			
4	PE	pattern	14.00	0.33	1.63	0.066	0.03
		Error(pattern)	770.00	0.21			
	RT	pattern	7.17	1.07	1.28	0.261	0.05
		Error(pattern)	186.45	0.84			
5	PE	pattern	14.00	0.22	2.25	0.006	0.06
		Error(pattern)	518.00	0.10			
	RT	pattern	8.84	1.30	5.54	0.000	0.14
		Error(pattern)	309.46	0.23			

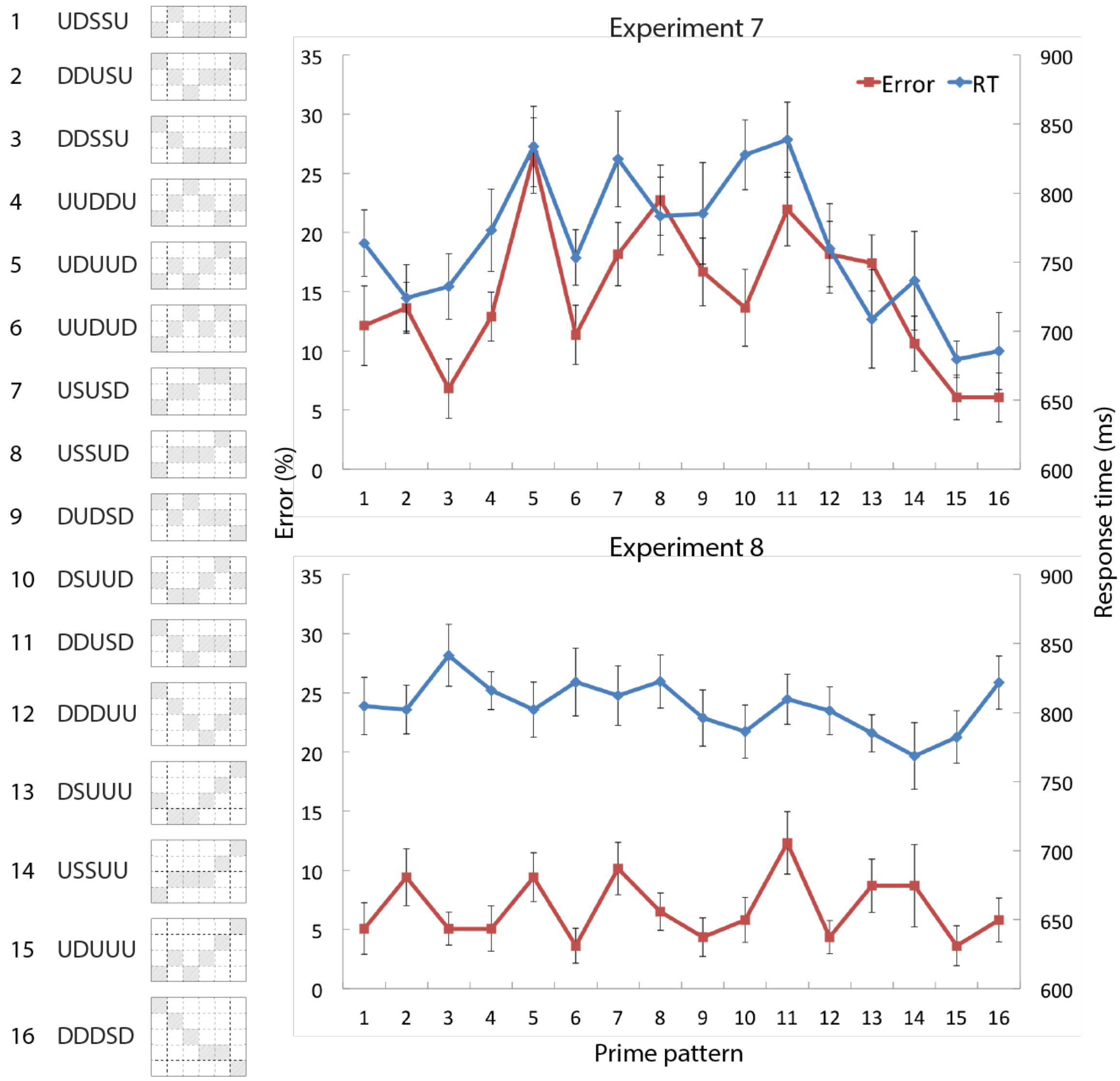
5.2 Experiment 6: Mean PE and RT in related conditions plotted as a function of standard pattern



ANOVA table – the effects of standard pattern on responses in related trials

Experiment	DV	Source	df	*MS*	*F*	sig.	η_p^2
6	PE	pattern	9.00	0.42	4.24	0.000	0.11
		Error(pattern)	315.00	0.10			
	RT	pattern	5.49	0.77	3.32	0.005	0.10
		Error(pattern)	164.60	0.23			

5.3 Experiments 7 and 8: Mean PE and RT in related conditions plotted as a function of prime pattern



ANOVA table – the effects of prime pattern on responses in related trials

Experiment	DV	Source	df	*MS*	*F*	sig.	η_p^2
7	RT	pattern	7.86	0.19	3.60	0.001	0.15
		Error(pattern)	165.09	0.05			
	PE	pattern	15.00	0.25	4.73	0.000	0.18
		Error(pattern)	315.00	0.05			
8	RT	pattern	7.60	0.04	0.94	0.481	0.04
		Error(pattern)	330.00	0.02			
	PE	pattern	8.05	0.11	1.40	0.198	0.06
		Error(pattern)	176.99	0.08			

References

Aarden, B. (2003). *Dynamic melodic expectancy* (Unpublished doctoral dissertation). Ohio State University, Columbus.

Aksentijevic, A., Elliott, M. A., & Barber, P. (2001). Dynamics of perceptual grouping: Similarities in the organization of visual and auditory groups. *Visual Cognition*, *8*(3), 349–358.

Aksentijevic, A., & Gibson, K. (2012a). Complexity equals change. *Cognitive Systems Research*, *15-16*, 1–16.

Aksentijevic, A., & Gibson, K. (2012b). Psychological complexity and the cost of information processing. *Theory & Psychology*, *22*(5), 572–590.

Alivisatos, B., & Petrides, M. (1997). Functional activation of the human brain during mental rotation. *Neuropsychologia*, *35*, 111–118.

Arnott, S. R., Binns, M. A., Grady, C. L., & Alain, C. (2004). Assessing the auditory dual-pathway model in humans. *NeuroImage*, *22*, 401–408.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, *61*(3), 183–93.

Attneave, F., & Olson, R. K. (1971). Pitch as a medium: A new approach to psychophysical scaling. *The American Journal of Psychology*, *84*(2), 147–166.

Balch, W. R. (1981). The role of symmetry in the good continuation ratings of two-part tonal melodies. *Perception & Psychophysics*, *29*(1), 47–55.

Balch, W. R., & Muscatelli, D. L. (1986). The interaction of modality condition and presentation rate in short-term contour recognition. *Perception & Psychophysics*, *40*(5), 351–358.

Barrett, D. J. K., & Hall, D. A. (2006). Response preferences for "what" and "where" in human non-primary auditory cortex. *NeuroImage*, *32*(2), 968–977.

Barrett, H. C., & Kurzban, R. (2006). Modularity in cognition: framing the debate. *Psychological Review*, *113*(3), 628–647.

Bartlett, J. C., & Dowling, W. J. (1980). Recognition of transposed melodies: A key-distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(3), 501–515.

Bartels, A., & Zeki, S. (2000). The architecture of the colour centre in the human visual brain: new results and a review. *The European Journal of Neuroscience*, *12*(1), 172–193.

Bernstein, I. H., & Edelstein, B. A. (1971). Effects of some variations in auditory input upon visual choice reaction time. *Journal of Experimental Psychology*, *87*(2), 241–247.

Berz, W. L. (1995). Working Memory in Music: A Theoretical Model. *Music Perception*, *12*(3), 353–364.

Billig, A., & Müllensiefen, D. (2012). Comparing Models of Melodic Contour in Music and Speech. In E. Cambouropoulos, C. Tsougras, P. Mavromatis, & K. Pastiadis (Eds.), *Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music* (pp. 95–96). Thessaloniki, Greece.

Bizley, J. K., & Cohen, Y. E. (2014). The what, where and how of auditory-object perception. *Nature Reviews Neuroscience*, *14*(10), 693–707.

Boltz, M. G. (1991). Some structural determinants of melody recall. *Memory & Cognition*, *19*(3), 239–251.

Boltz, M. G., & Jones, M. R. (1986). Does rule recursion make melodies easier to reproduce? If not, what does? *Cognitive Psychology*, *18*, 389–431.

Boltz, M. G., Marshburn, E., & Jones, M. R. (1985). Serial-pattern structure and temporal-order recognition. *Perception & Psychophysics*, *37*(3), 209–217.

Bregman, A. S. (1978). The formation of auditory streams. In J. Requin (Ed.), *Attention and Performance VII*. Hillsdale, NJ: Erlbaum.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press.

Brunel, L., Labeye, E., Lesourd, M., & Versace, R. (2009). The sensory nature of episodic memory: sensory priming effects due to memory trace activation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 1081–1088.

Buffart, H., Leeuwenberg, E. L. J., & Restle, F. (1981). Coding theory of visual pattern completion. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(2), 241–274.

Buhusi, C. V, & Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nature Reviews. Neuroscience*, *6*(10), 755–765.

Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews*, *36*, 96–107.

Buonomano, D. V, & Karmarkar, U. R. (2002). How do we tell time? *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, *8*(1), 42–51.

Burns, E. M., & Ward, W. D. (2013). Intervals, scales, and tuning. In D. Deutsch (Ed.), *The Psychology of Music* (3rd ed.). New York: Academic Press.

Burt, P., & Sperling, G. (1981). Time, distance, and feature trade-offs in visual apparent motion. *Psychological Review*, *88*, 171–195.

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, *51*(13), 1484–1525.

Carter, M. D., Hough, M. S., Stuart, A., & Rastatter, M. P. (2011). The effects of inter-stimulus interval and prime modality in a semantic priming task. *Aphasiology*, *25*(6-7), 761–773.

Casasanto, D. (2010). Space for Thinking. In V. Evans & P. Chilton (Eds.), *Language, Cognition, and Space: State of the Art and New Directions* (pp. 453–478). London: Equinox Publishing.

Chater, N. (1996). Reconciling simplicity and likelihood principles in perceptual organization. *Psychological Review*, *103*(3), 566–581.

Chater, N. (1999). The Search for Simplicity: A Fundamental Cognitive Principle? *The Quarterly Journal of Experimental Psychology Section A*, *52*(2), 273–302.

Cheveigné, A. de. (2004). Pitch perception models. In C. Plack & A. J. Oxenham (Eds.), *Pitch*. New York: Springer Verlag.

Cho, Y. S., & Proctor, R. W. (2003). Stimulus and response representations underlying orthogonal stimulus – response compatibility effects. *Psychonomic Bulletin & Review*, *10*(1), 45–73.

Cohen Kadosh, R., & Henik, A. (2006). A common representation for semantic and physical properties: A cognitive-anatomical approach. *Experimental Psychology*, *53*(2), 87–94.

Cohen Kadosh, R., Lammertyn, J., & Izard, V. (2008). Are numbers special? An overview of chronometric, neuroimaging, developmental and comparative studies of magnitude representation. *Progress in Neurobiology*, *84*(2), 132–147.

Collard, R., & Povel, D.-J. (1982). Theory of serial pattern production: tree traversals. *Psychological Review*, *89*(6), 693–707.

Collier, G. L., & Logan, G. (2000). Modality differences in short-term memory for rhythms. *Memory & Cognition*, *28*(4), 529–538.

Correa, Á., Lupiáñez, J., Madrid, E., & Tudela, P. (2006). Temporal attention enhances early visual processing: A review and new evidence from event-related potentials. *Brain Research*, *1076*(1), 116–128.

Creelman, C. D. (1962). Human Discrimination of Auditory Duration. *The Journal of the Acoustical Society of America*, *34*(5), 582.

Creem, S. H., & Proffitt, D. R. (2001). Defining the cortical visual systems: "what", "where", and "how". *Acta Psychologica*, *107*, 43–68.

Cross, I. (1985). Music and change: On the establishment of rules. In P. Howell, I. Cross, & R. West (Eds.), *Musical Structure and Cognition* (pp. 1–19). London: Academic Press Inc. (London) Ltd.

Cuddy, L. L., & Cohen, A. J. (1976). Recognition of transposed melodic sequences. *Quarterly Journal of Experimental Psychology*, *28*(2), 255–270.

Cuddy, L. L., & Lyons, H. I. (1981). Musical pattern recognition: a comparison of listening to and studying tonal structures and tonal ambiguities. *Psychomusicology*, *I*(2), 15–33.

Cupchik, G. C., Phillips, K., & Hill, D. S. (2001). Shared processes in spatial rotation and musical permutation. *Brain and Cognition*, *46*(3), 373–382.

De la Mothe, L. A., Blumell, S., Kajikawa, Y., & Hackett, T. A. (2006). Cortical connections of the auditory cortex in marmoset monkeys: core and medial belt regions. *Journal of Comparative Neurology*, *496*, 27–71.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371–396.

Deutsch, D. (1980). The processing of structured and unstructured tonal sequences. *Perception & Psychophysics*, *28*(5), 381–389.

Deutsch, D., & Feroe, J. (1981). The internal representation of pitch sequences in tonal music. *Psychological Review*, *88*(6), 503–522.

DeYoe, E. A., & Van Essen, D. C. (1988). Concurrent processing streams in monkey visual cortex. *Trends in Neurosciences*, *11*, 219–226.

Dienes, Z., & Longuet-Higgins, H. C. (2004). Can musical transformations be implicitly learned? *Cognitive Science*, *28*(4), 531–558.

Dowling, W. J. (1971). Recognition of inversions of melodies and melodic contours. *Perception & Psychophysics*, *9*(3B), 348–349.

Dowling, W. J. (1972). Recognition of melodic transformations: Inversion, retrograde, and retrograde inversion. *Perception & Psychophysics*, *12*(5), 417–421.

Dowling, W. J. (1973). The perception if interleaved melodies. *Cognitive Psychology*, *5*, 322–337.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, *85*(4), 341–354.

Dowling, W. J., & Bartlett, J. C. (1981). The importance of interval information in long-term memory for melodies. *Psychomusicology*, *1*(1), 30–49.

Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *The Journal of the Acoustical Society of America*, *49*(2), 524–531.

Dowling, W. J., Lung, K. M.-T., & Herrbold, S. (1987). Aiming attention in pitch and time in the perception of interleaved melodies. *Perception & Psychophysics*, *41*(6), 642–656.

Dyson, M. C., & Watkins, A. J. (1984). A figural approach to the role of melodic contour in melody recognition. *Perception & Psychophysics*, *35*(5), 477–88.

Edworthy, J. (1985). Interval and contour in melody processing. *Music Perception*, *2*(3), 375–388.

Eerola, T., Jarvinen, T., Louhivuori, J., & Toiviainen, P. (2001). Statistical Features and Perceived Similarity of Folk Melodies. *Music Perception*, *18*(3), 275–296.

Eitan, Z., & Granot, R. Y. (2006). How music moves: musical parameters and listeners' images of motion. *Music Perception*, *23*(3), 221–247.

Eitan, Z., & Timmers, R. (2010). Beethoven's last piano sonata and those who follow crocodiles: cross-domain mappings of auditory pitch in a musical context. *Cognition*, *114*(3), 405–422.

Ellis, R. J., & Jones, M. R. (2009). The role of accent salience and joint accent structure in meter perception. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(1), 264–80.

Essens, P. J. (1986). Hierarchical organization of temporal patterns. *Perception & Psychophysics*, *40*(2), 69–73.

Evans, K. K., & Treisman, A. M. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, *10*(1:6), 1–12.

Falk, R., & Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review*, *104*(2), 301–318.

Fischer, M. H. (2003). Spatial representations in number processing - evidence from a pointing task. *Visual Cognition*, *10*(4), 493–509.

Foster, N. E. V, Halpern, A. R., & Zatorre, R. J. (2013). Common parietal activation in musical mental transformations across pitch and time. *NeuroImage*, *75*, 27–35.

Foster, N. E. V, & Zatorre, R. J. (2010a). A role for the intraparietal sulcus in transforming musical pitch information. *Cerebral Cortex*, *20*(6), 1350–1359.

Foster, N. E. V, & Zatorre, R. J. (2010b). Cortical structure predicts success in performing musical transformation judgments. *NeuroImage*, *53*(1), 26–36.

Fountain, S. B., & Rowan, J. D. (1995). Coding of hierarchical versus linear pattern structure in rats and humans. *Journal of Experimental Psychology: Animal Behavior Processes*, *21*(3), 187–202.

Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The Psycholgy of Music* (pp. 149–180). New York: Academic Press.

Fraisse, P. (1984). Perception and estimation of time. *Annual Review of Psychology*, *35*, 1–36.

Freedman, E. G. (1999). The Role of Diatonicism in the Abstraction and Representation of Contour and Interval Information. *Music Perception: An Interdisciplinary Journal*, *16*(3), 365–387.

Frey, S. H., Vinton, D., Norlund, R., & Grafton, S. T. (2005). Cortical topography of human anterior intraparietal cortex active during visually guided grasping. *Cognitive Brain Research*, *23*, 397–406.

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., & Pantev, C. (2004). Musical training enhances automatic encoding of melodic contour and interval structure. *Journal of Cognitive Neuroscience*, *16*(6), 1010–1021.

Gallace, A., & Spence, C. (2011). To what extent do Gestalt grouping principles influence tactile perception? *Psychological Bulletin*, *137*(4), 538–61.

Garner, W. R., & Gottwald, R. L. (1968). The perception and learning of temporal patterns. *The Quarterly Journal of Experimental Psychology*, *20*(2), 97–109.

Gault, R. H., & Goodfellow, L. D. (1938). An Empirical Comparison of Audition, Vision, and Touch in the Discrimination of Temporal Patterns and Ability to Reproduce them. *The Journal of General Psychology*, *18*(1), 41–47.

Gevers, W., Reynvoet, B., & Fias, W. (2003). The mental representation of ordinal sequences is spatially organized. *Cognition*, *87*, B87–B95.

Ginsburg, V., & Gevers, W. (2015). Spatial coding of ordinal information in short- and long-term memory. *Frontiers in Human Neuroscience*, *9*(5), 1–10.

Goldstein, J. L. (1973). An optimum processor theory for the central formation of the pitch of complex tones. *Journal of the Acoustical Society of America*, *54*, 1496–1516.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(I), 20–25.

Gordon, E. E. (1975). *Learning theory, patterns, and music*. Buffalo, NY: Tometic Associates.

Gosselin, N., Jolicoeur, P., & Peretz, I. (2009). Impaired memory for pitch in congenital amusia. *Annals of the New York Academy of Sciences*, *1169*, 270–272.

Greenberg, G. Z., & Larkin, W. D. (1968). Frequency-response characteristic of auditory observers detecting signals of a single frequency in noise: the probe-signal method. *Journal of the Acoustical Society of America*, *44*(6), 1513–1523.

Grefkes, C., & Fink, G. R. (2005). The functional organization of the intraparietal sulcus in humans and monkeys. *Journal of Anatomy*, *207*(1), 3–17.

Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nature Reviews Neuroscience*, *5*(November), 887–892.

Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annual Review of Neuroscience*, *27*, 649–677.

Grondin, S., & Rousseau, R. (1991). Judging the relative duration of multimodal short empty time intervals. *Perception & Psychophysics*, *49*(3), 245–256.

Hackett, T. A., Preuss, T. M., & Kaas, J. H. (2001). Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *Journal of Comparative Neurology*, *441*, 197–222.

Hackett, T. A., Stepniewska, I., & Kaas, J. H. (1998a). Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *Journal of Comparative Neurology*, *394*, 475–495.

Hackett, T. A., Stepniewska, I., & Kaas, J. H. (1998b). Thalamocortical connections of the parabelt auditory cortex in macaque monkeys. *Journal of Comparative Neurology*, *400*, 271–286.

Hahn, U. (2014). Similarity. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(3), 271–280.

Hahn, U., Chater, N., & Richardson, L. B. (2003). Similarity as transformation. *Cognition*, *87*(1), 1–32.

Halpern, A. R., Bartlett, J. C., & Dowling, W. J. (1998). Perception of mode, rhythm, and contour in unfamiliar melodies: Effects of age and experience. *Music Perception: An Interdisciplinary Journal*, *15*(4), 335–355.

Handel, S. (1993). *Listening: An Introduction to the Perception of Auditory Events*. Cambridge, MA: MIT Press.

Harris, I. M., Egan, G. F., Sonkilla, C., Tochon-Danguy, H. J., Paxinos, G., & Watson, J. D. (2000). Selective right parietal lobe activation during mental rotation: a parametric PET study. *Brain*, *123*(1), 65–73.

Helmholtz, H. L. F. von. (1954). *On the Sensations of Tone as a Physiological Basis for the Theory of Music*. (A. J. Ellis, Ed.). New York: Dover Publishing.

Helmholtz, H. L. F. von. (1962). *Treatise on Physiological Optics (J. P. C. Southall, Trans.)*. New York: Dover.

Hochberg, J. E., & McAlister, E. (1953). A quantitative approach to figural "goodness." *Journal of Experimental Psychology*, *46*(5), 361–364.

Hochberg, J. E., & Silverstein, A. (1956). A quantitative index of stimulus-similarity: Proximity versus differences in brightness. *American Journal of Psychology*, *69*, 456–458.

Hodgetts, C. J., & Hahn, U. (2012). Similarity-based asymmetries in perceptual matching. *Acta Psychologica*, *139*(2), 291–299.

Howard, J. H., O'Toole, A. J., Parasuraman, R., & Bennett, K. B. (1984). Pattern-directed attention in uncertain-frequency detection. *Perception & Psychophysics*, *35*(3), 256–264.

Howell, D. C. (2002). *Statistical Methods for Psychology* (5th ed.). Belmont, CA: Duxbury Press.

Hubbard, T. L., & Bharucha, J. J. (1988). Judged displacement in apparent vertical and horizontal motion. *Perception & Psychophysics*, *44*(3), 211–221.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, *195*, 215–243.

Ischebeck, A., Zamarian, L., Siedentopf, C., Koppelstatter, F., Benke, T., Felber, S., & Delazer, M. (2006). How specifically do we learn? Imaging the learning of multiplication and subtraction. *NeuroImage*, *30*, 1365–1375.

Ishihara, M., Keller, P. E., Rossetti, Y., & Prinz, W. (2008). Horizontal spatial representations of time: evidence for the STEARC effect. *Cortex*, *44*(4), 454–461.

Iversen, J. R., Patel, A., Nicodemus, B., & Emmorey, K. (2015). Synchronization to auditory and visual rhythms in hearing and deaf individuals. *Cognition*, *134*, 232–244.

Johnson, R. C., McClearn, G. E., Yuen, S., Nagoshi, C. T., Ahern, F. M., & Cole, R. E. (1985). Galton's data a century later. *American Psychologist*, *40*(8), 875–892.

Johnsrude, I. S., Penhune, V. B., & Zatorre, R. J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain*, *123*, 155–163.

Johnston, H. M., & Jones, M. R. (2006). Higher order pattern structure influences auditory representational momentum. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(1), 2–17.

Jones, M. R. (1974). Cognitive representations of serial patterns. In B. Kantowitz (Ed.), *Human Information Processing: Tutorials in Performance Cognition*. Potomac, MD: Erlbaum.

Jones, M. R. (1976a). Levels of structure in the reconstruction of temporal and spatial serial patterns. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(4), 475–488.

Jones, M. R. (1976b). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, *83*(5), 323–355.

Jones, M. R. (1978). Auditory Patterns: The Perceiving organism. In E. C. Carterette & M. P. Friedman (Eds.), *The Handbook of Perception (Vol. VIII)*. New York: Academic Press.

Jones, M. R. (1981). A tutorial on some issues and methods in serial pattern research. *Perception & Psychophysics*, *30*(5), 492–504.

Jones, M. R. (1987). Dynamic pattern structure in music: Recent theory and research. *Perception & Psychophysics*, *41*(6), 621–34.

Jones, M. R., & Boltz, M. G. (1989). Dynamic attending and responses to time. *Psychological Review*, *96*(3), 459–491.

Jones, M. R. (1990). Learning and the development of expectancies. *Psychomusicology*, *9*(2), 193–228.

Jones, M. R., Johnston, H. M., & Puente, J. (2006). Effects of auditory pattern structure on anticipatory and reactive attending. *Cognitive Psychology*, *53*(1), 59–96.

Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal Aspects of Stimulus-Driven Attending in Dynamic Arrays. *Psychological Science*, *13*(4), 313–319.

Jones, M. R., Summerell, L., & Marshburn, E. (1987). Recognizing melodies: A dynamic interpretation. *The Quarterly Journal of Experimental Psychology Section A*, *39*(1), 89-121.

Jones, M. R., & Zamostny, K. P. (1975). Memory and rule structure in the prediction of serial patterns. *Journal of Experimental Psychology: Human Learning and Memory*, *1*(3), 295–306.

Julesz, B. (1971). *Foundation of Cylopean Perception*. Cambridge, Mass: MIT Press.

Justus, T. C., & Bharucha, J. J. (2002). Music perception and cognition. In S. Yantis & H. Pashler (Eds.), *Stevens' Handbook of Experimental Psychology, Volume 1: Sensation and Perception (Third Edition)* (pp. 453–492). New York: Wiley.

Kaas, J. H., & Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(22), 11793–11799.

Kaas, J. H., Hackett, T. A., & Tramo, M. J. (1999). Auditory processing in primate cerebral cortex. *Current Opinion in Neurobiology*, *9*, 164–170.

Koch, I., & Hoffmann, J. (2000). The role of stimulus-based and response-based spatial information in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(4), 863–882.

Koffka, K. (1935). *Principles of Gestalt Psychology*. London: Lund Humphries.

Köhler, W. (1929). *Gestalt Psychology*. New York: H. Liveright.

Kolmogorov, A. N. (1968). Three approaches to the quantitative definition of information. *International Journal of Computer Mathematics*, *2*(1-4), 157–168.

Kong, J., Wang, C., Kwong, K., Vangel, M., Chua, E., & Gollub, R. (2005). The neural substrate of arithmetic operations and procedure complexity. *Cognitive Brain Research*, *2*, 397–405.

Kosslyn, S. M. (1980). *Image and Mind*. Cambridge, MA: Harvard University Press.

Kotovsky, K., & Simon, H. A. (1973). Empirical tests of a theory of human acquisition for sequential of concepts. *Cognitive Psychology*, *4*, 399–424.

Koukkari, W. L., & Sothern, R. B. (2006). *Introducing Biological Rhythms: A Primer on the Temporal Organization of Life, with Implications for Health, Society, Reproduction and the Natural Environment*. New York: Springer.

Krumhansl, C. L. (2000). Rhythm and pitch in music cognition. *Psychological Bulletin*, *126*(1), 159–179.

Krumhansl, C. L., Sandell, G. J., & Sergeant, D. C. (1987). The Perception of Tone Hierarchies and Mirror Forms in Twelve-Tone Serial Music. *Music Perception: An Interdisciplinary Journal*, *5*(1), 31–77.

Kubovy, M. (1988). Should we resist the seductiveness of the space:time::vision:audition analogy? *Journal of Experimental Psychology: Human Perception and Performance*, *14*(2), 318–320.

Kubovy, M., & Valkenburg, D. van. (2001). Auditory and visual objects. *Cognition*, *80*(1-2), 97–126.

Kuhn, G., & Dienes, Z. (2005). Implicit learning of nonlocal musical rules: implicitly learning more than chunks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1417–1432.

Kundey, S. M. A., De Los Reyes, A., Rowan, J. D., Lee, B., Delise, J., Molina, S., & Cogdill, L. (2013). Involvement of working memory in college students' sequential pattern learning and performance. *Learning and Motivation*, *44*(2), 114–126.

Kundey, S. M. A., & Rowan, J. D. (2014). Hierarchical organization in serial pattern learning. *Learning and Motivation*, *46*, 60–68.

Lakens, D., Semin, G. R., & Garrido, M. V. (2011). The sound of time: cross-modal convergence in the spatial structuring of time. *Consciousness and Cognition*, *20*(2), 437–443.

Lakoff, G., & Johnson, M. (1980). The Metaphorical Structure of the Human Conceptual System. *Cognitive Science*, *4*(2), 195–208.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: how people track time-varying events. *Psychological Review*, *106*(1), 119–159.

Lee, Y.-S., Janata, P., Frost, C., Hanke, M., & Granger, R. (2011). Investigation of melodic contour processing in the brain using multivariate pattern-based fMRI. *NeuroImage*, *57*(1), 293–300.

Lee, Y.-S., Janata, P., Frost, C., Martinez, Z., & Granger, R. (2015). Melody recognition revisited: Influence of melodic Gestalt on the encoding of relational pitch information. *Psychonomic Bulletin & Review*, *22*(1), 163-169.

Leeuwenberg, E. L. J. (1968). *Structural Information of Visual Patterns: An Efficient Coding System in Perception*. The Hague: Mouton.

Leeuwenberg, E. L. J. (1969). Quantitative specification of information in sequential patterns. *Psychological Review*, *76*(2), 216–220.

Lerdahl, F., & Jackendoff, R. (1983). *The Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.

Levarie, S., & Levy, E. (1981). *Tone: A Study in Music Acoustics*. Greenwood Press.

Li, M., & Vitányi, P. M. B. (1997). *Introduction to Kolmogorov Complexity and its Applications* (2nd ed.). New York: Springer-Verlag.

Lidji, P., Kolinsky, R., Lochy, A., & Morais, J. (2007). Spatial associations for musical stimuli: a piano in the head? *Journal of Experimental Psychology: Human Perception and Performance*, *33*(5), 1189–1207.

Liégeois-Chauvel, C., Peretz, I., Babaï, M., Laguitton, V., & Chauvel, P. (1998). Contribution of different cortical areas in the temporal lobes to music processing. *Brain*, *121*, 1853–1867.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*(4), 476–490.

Luce, R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization*. New York: Oxford University Press.

MacKay, D. M. (1950). XXIV. Quantal aspects of scientific information. *Philosophical Magazine*, *41*(314), 289–311.

Marmel, F., Tillmann, B., & Delbé, C. (2010). Priming in melody perception: tracking down the strength of cognitive expectations. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(4), 1016–1028.

Mazzola, G. (2013). *The Topos of Music: Geometric Logic of Concepts, Theory, and Performance*. Basel: Birkhäuser.

McConnell, J., & Quinn, J. G. (2004). Complexity factors in visuo-spatial working memory. *Memory*, *12*(3), 338–350.

McDermott, J. H., Lehr, A. J., & Oxenham, A. J. (2008). Is relative pitch specific to pitch? *Psychological Science*, *19*(12), 1263–1271.

McLachlan, N. M., Greco, L. J., Toner, E. C., & Wilson, S. J. (2010). Using spatial manipulation to examine interactions between visual and auditory encoding of pitch and time. *Frontiers in Psychology*, *1*(December), 1–14.

McLachlan, N. M., & Wilson, S. J. (2010). The central role of recognition in auditory perception: A neurobiological model. *Psychological Review*, *117*(1), 175–96.

Meiser, T. (2011). Much Pain, Little Gain? Paradigm-Specific Models and Methods in Experimental Psychology. *Perspectives on Psychological Science*, *6*(2), 183–191.

Miles, L. K., Nind, L. K., & Macrae, C. N. (2010). Moving through time. *Psychological Science, 21*(2), 222–223.

Miyazaki, K. (2004). Recognition of transposed melodies by absolute-pitch possessors. *Japanese Psychological Research*, *46*(4), 270–282.

Moerel, M., De Martino, F., & Formisano, E. (2014). An anatomical and functional topography of human auditory cortical areas. *Frontiers in Neuroscience*, *8*(July), 1–14.

Morgan, R. P. (1998). Symmetrical form and common-practice tonality. *Music Theory Spectrum*, *20*(1), 1–47.

Mudd, S. A. (1963). Spatial stereotypes of four dimensions of pure tone. *Journal of Experimental Psychology*, *66*(4), 347–352.

Mukari, S. Z., Umat, C., & Othman, N. I. (2010). Effects of age and working memory capacity on pitch pattern sequence test and dichotic listening. *Audiology & Neurotology*, *15*(5), 303–310.

Müllensiefen, D., & Wiggins, G. (2011). Polynomial functions as a representation of melodic phrase contour. In A. Schneider & A. von Ruschowski (Eds.), *Systematic Musicology: Empirical and Theoretical Studies* (pp. 63–88). Frankfurt: Peter Lang.

Nobre, A. C. (2001). Orienting attention to instants in time. *Neuropsychologia*, *39*, 1317–1328.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57.

O'Leary, A., & Rhodes, G. (1984). Cross-modal effects on visual and auditory object perception. *Perception & Psychophysics*, *35*(6), 565–569.

O'Shaughnessy, D. (1987). *Speech Communication: Human and Machine*. Reading, Massachusetts: Addison-Wesley.

Olivers, C. N. L., & van der Helm, P. A. (1998). Symmetry and selective attention: A dissociation between effortless perception and serial search. *Perception & Psychophysics*, *60*(7), 1101–1116.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716.

Ortony, A. (1979). Beyond literal similarity. *Psychological Review*, *86*(3), 161–180.

Overath, T., Cusack, R., Kumar, S., Kriegstein, K. von, Warren, J. D., Grube, M., Carlyon, R. P., & Griffiths, T. D. (2007). An information theoretic characterisation of auditory encoding. *PLoS Biol*, *5*(11), e288.

Pachella, R. G. (1974). The interpretation of reaction time in information-processing research. In B. H. Kantowitz (Ed.), *Human Information Processing: Tutorials in Performance Cognition* (pp. 41–82). Hillsdale, NJ: Lawrence Erlbaum Associates Ltd.

Palmer, S. E. (1983). The psychology of perceptual organization: A transformational approach. In J. Beck, B. Hope, & A. Rosenfeld (Eds.), *Human and Machine Vision* (pp. 269–339). New York: Academic Press.

Palmer, S. E., & Hemenway, K. (1978). Orientation and symmetry: effects of multiple, rotational, and near symmetries. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(4), 691–702.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, *36*(4), 767–776.

Pedersen, P. (1965). The mel scale. *Journal of Music Theory*, *9*(2), 295–308.

Peretz, I. (1990). Processing of local and global musical information by unilateral brain-damaged patients. *Brain*, *113*(4), 1185–1205.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, *6*(7), 688–691.

Petkov, C. I., Kayser, C., Augath, M., & Logothetis, N. K. (2006). Functional imaging reveals numerous fields in the monkey auditory cortex. *PLoS Biology*, *4*(7), e215.

Pietsch, S., & Jansen, P. (2012). Different mental rotation performance in students of music, sport and education. *Learning and Individual Differences*, *22*(1), 159–163.

Pomerantz, J. R., & Lockhead, G. R. (1991). Perception of structure: an overview. In G. R. Lockhead & J. R. Pomerantz (Eds.), *The Perception of Structure: Essays in Honor of Wendell R. Garner* (pp. 1–20). Washington, DC: American Psychological Association.

Pomerantz, J. R., & Portillo, M. C. (2011). Grouping and emergent features in vision: Toward a theory of basic Gestalts. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(5), 1331–1349.

Posner, M. I., Nissen, M. J., & Ogden, W. C. (1978). Attended and unattended processing modes. In H. L. Pick & I. J. Saltzman (Eds.), *Modes of perceiving and processing modes*. Hillsdale, NJ: Erlbaum.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology*, *109*(2), 160–174.

Posner, M. I., & Snyder, C. R. R. (1975a). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium*. Potomac, MD: Erlbaum.

Posner, M. I., & Snyder, C. R. R. (1975b). Facilitation and inhibition in the processing of signals. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and Performance V*. London: Academic Press.

Povel, D.-J. (1981). Internal representation of simple temporal patterns. *Journal of Experimental Psychology. Human Perception and Performance*, *7*(1), 3–18.

Povel, D.-J. (1984). A theoretical framework for rhythm perception. *Psychological Research*, *45*, 315–337.

Povel, D.-J., & Essens, P. J. (1985). Perception of temporal patterns. *Music Perception*, *2*(4), 411–440.

Pratt, C. C. (1930). The spatial character of high and low tones. *Journal of Experimental Psychology*, *13*, 278–285.

Prince, J. B., Schmuckler, M. A., & Thompson, W. F. (2009). Cross-modal melodic contour similarity. *Canadian Acoustics*, *37*(1), 35–49.

Quinlan, P. T., & Wilton, R. N. (1998). Grouping by proximity or similarity? Competition between the Gestalt principles in vision. *Perception*, *27*, 417–430.

Quinn, I. (1999). The Combinatorial Model of Pitch Contour. *Music Perception*, *16*(4), 439–456.

Rauschecker, J. P. (2013). Processing streams in auditory cortex. In Y. E. Cohen, A. Popper, & R. R. Fay (Eds.), *Neural Correlates of Auditory Cognition* (pp. 7–43). New York: Springer-Verlag.

Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(22), 11800–11806.

Reiner, T. W. (2011). Pitch-distance and contour complexity in the recognition of short melodies. *Journal of Scientific Psychology*, (September), 27–36.

Restle, F. (1970). Theory of serial pattern learning: Structural trees. *Psychological Review*, *77*(6), 481–495.

Restle, F. (1973). Serial pattern learning: higher order transitions. *Journal of Experimental Psychology*, *99*(1), 61–69.

Restle, F. (1976). Structural ambiguity in serial pattern learning. *Cognitive Psychology*, *8*, 357–381.

Restle, F., & Brown, E. R. (1970). Serial pattern learning. *Journal of Experimental Psychology*, *83*(1), 120–125.

Ross, D. a, Gore, J. C., & Marks, L. E. (2005). Absolute pitch: music and beyond. *Epilepsy & Behavior*, *7*(4), 578–601.

Rousseau, R., Poirier, J., & Lemyre, L. (1983). Duration discrimination of empty time intervals marked by intermodal pulses. *Perception & Psychophysics*, *34*(6), 541–548.

Royer, F. L. (1981). Detection of symmetry. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(6), 1186–1210.

Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C., & Butterworth, B. (2006). Spatial representation of pitch height: The SMARC effect. *Cognition*, *99*(2), 113–129.

Schmithorst, V. J. (2005). Separate cortical networks involved in music perception: preliminary functional MRI evidence for modularity of music processing. *NeuroImage*, *25*(2), 444–451.

Schmuckler, M. A. (1999). Testing Models of Melodic Contour Similarity. *Music Perception: An Interdisciplinary Journal*, *16*(3), 295–326.

Schmuckler, M. A. (2009). Components of melodic processing. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford Handbook of Music Psychology*. Oxford: Oxford University Press.

Schoenberg, A. (1975). *Style and Idea (translations by Leo Black)*. (L. Stein, Ed.). London: Faber & Faber Ltd.

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, *80*(1-2), 1–46.

Schulze, K., Dowling, W. J., & Tillmann, B. (2012). Working memory for tonal and atonal sequences during a forward and a backward recognition task. *Music Perception*, *29*(3), 255–267.

Schuppert, M., Münte, T. F., Wieringa, B. M., & Altenmüller, E. (2000). Receptive amusia: evidence for cross-hemispheric neural networks underlying music processing strategies. *Brain : A Journal of Neurology*, *123 Pt 3*, 546–59.

Shaki, S., & Fischer, M. H. (2008). Reading space into numbers - a cross-linguistic comparison of the SNARC effect. *Cognition*, *108*(2), 590–599.

Shamma, S. A. (2008). On the emergence and awareness of auditory objects. *PLoS Biology*, *6*(6), e155.

Shamma, S. A., & Klein, D. (2000). The case of the missing pitch templates: How harmonic templates emerge in the early auditory system. *Journal of the Acoustical Society of America*, *107*, 2631–2644.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*, 623–656.

Shepard, R. N. (1957). Stimulus and response generalization: a stochastic model relating generalization to distance in psychological space. *Psychometrika*, *22*(4), 325–345.

Shepard, R. N. (1964). Circularity in judgments of relative pitch. *The Journal of the Acoustical Society of America*, *36*(12), 2346–2353.

Shepard, R. N. (1982). Geometrical approximations to the structure of musical pitch. *Psychological Review*, *89*(4), 305–333.

Shepard, R. N. (1987). Towards a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701–703.

Shelton, J., & Kumar, G. P. (2010). Comparison between Auditory and Visual Simple Reaction Times. *Neuroscience & Medicine*, *1*, 30–32.

Sidtis, J. J. (1980). On the nature of the cortical function underlying right hemisphere auditory perception. *Neuropsychologia*, *18*(3), 321–330.

Simon, H. A. (1972). Complexity and the representation of patterned sequences of symbols. *Psychological Review*, *79*(5), 369–382.

Simon, H. A., & Kotovsky, K. (1963). Human acquisition of concepts for sequential patterns. *Psychological Review*, *70*(6), 534–546.

Solomonoff, R. J. (1964). A formal theory of inductive inference, part 1 and part 2. *Information & Control*, *7*, 224–254.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*, 971–995.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103*(2684), 677–680.

Stewart, L., Overath, T., Warren, J. D., Foxton, J. M., & Griffiths, T. D. (2008). fMRI evidence for a cortical hierarchy of pitch pattern processing. *PloS ONE*, *3*(1), e1470.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*(9), 403–409.

Talavage, T. M., Sereno, M. I., Melcher, J. R., Ledden, P. J., Rosen, B. R., & Dale, A. M. (2004). Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. *Journal of Neurophysiology*, *91*(3), 1282–1296.

Tanner, W. P., & Norman, R. Z. (1954). The human use of information-II: Signal detection for the case of an unknown signal parameter. *Transactions of the IRE Professional Group on Information Theory (TIT)*, *4*, 222–227.

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, *21*, 233–282.

Terhardt, E. (1974). Pitch, consonance, and harmony. *Journal of the Acoustical Society of America*, *55*, 1061–1069.

Thomassen, J. M. (1982). Melodic accent: Experiments and a tentative model. *The Journal of the Acoustical Society of America*, *71*(6), 1596.

Thorpe, M., Ockelford, A., & Aksentijevic, A. (2012). An empirical exploration of the zygonic model of expectation in music. *Psychology of Music*, *40*(4), 429–470.

Tootell, R. B. H., & Hadjikhani, N. (2000). Attention - brains at work! *Nature Neuroscience*, *3*(3), 206–208.

Tootell, R. B. H., Reppas, J. B., Kwong, K. K., Malach, R., Born, R. T., Brady, T. J., … Belliveau, J. W. (1995). Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *Journal of Neuroscience*, *15*(4), 3215–3230.

Trainor, L. J., Desjardins, R. N., & Rockel, C. (1999). A comparison of contour and interval processing in musicians and nonmusicians using event-related potentials. *Australian Journal of Psychology*, *51*(3), 147–153.

Tramo, M. J., Shah, G. D., & Braida, L. D. (2002). Functional role of auditory cortex in frequency processing and pitch perception. *Journal of Neurophysiology*, *87*, 122–139.

Treder, M. S. (2010). Behind the looking-glass: a review on human symmetry perception. *Symmetry*, *2*(3), 1510–1543.

Trehub, S. E., Bull, D., & Thorpe, L. A. (1984). Infants' perception of melodies: The role of melodic contour. *Child Development*, *55*(3), 821–830.

Treisman, M. (1963). Temporal discrimination and the indifference interval: implications for a model of the "internal clock." *Psychological Monographs*, *77*, 1–31.

Turatto, M., Mazza, V., & Umiltà, C. (2005). Crossmodal object-based attention: auditory objects affect visual processing. *Cognition*, *96*(2), B55–64.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327–352.

Ungerleider, L. G., & Haxby, J. V. (1994). "What" and "where" in the human brain. *Current Opinion in Neurobiology*, *4*, 157–165.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of Visual Behavior* (pp. 549–586). Cambridge, MA: MIT Press.

Van der Helm, P. A. (2000). Simplicity versus likelihood in visual perception: from surprisals to precisals. *Psychological Bulletin*, *126*(5), 770–800.

Van der Helm, P. A., & Leeuwenberg, E. L. J. (1991). Accessibility, a criterion for regualrity and hierarchy in visual pattern codes. *Journal of Mathematical Psychology*, *35*, 151–213.

Van der Helm, P. A., & Leeuwenberg, E. L. J. (1996). Goodness of visual regularities: a nontransformational approach. *Psychological Review*, *103*(3), 429–456.

Van Essen, D. C., & Maunsell, J. H. R. (1983). Hierarchical organization and functional streams in the visual cortex. *Trends in Neurosciences*, *6*, 370–375.

Vitányi, P. M. B., & Li, M. (2000). Minumum description length induction, Bayesianism, and Kolmogorov complexity. *IEEE Transactions on Information Theory*, *46*, 446–464.

Vitz, P. C., & Todd, T. C. (1969). A coded element model of the perceptual processing of sequential stimuli. *Psychological Review*, *76*(5), 433–449.

Von Ehrenfels, C. (1890). Über "Gestaltqualitäten." *Vierteljahrschrift Für Wissenschaftliche Philosophie*, *14*, 249–292.

Von Ehrenfels, C. (1937). On Gestalt-qualities. *Psychological Review*, *44*(6), 521–524.

Vu, K.-P. L., Proctor, R. W., & Pick, D. F. (2000). Vertical versus horizontal spatial compatibility: Right-left prevalence with bimanual responses. *Psychological Research*, *64*, 25–40.

Wagemans, J. (1995). Detection of visual symmetries. *Spatial Vision*, *9*(1), 9–32.

Wagemans, J. (1997). Characteristics and models of human symmetry detection. *Trends in Cognitive Sciences*, *1*(9), 346–352.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. a, Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin*, *138*(6), 1172–1217.

Wallace, J. M., & Scott-Samuel, N. E. (2007). Spatial versus temporal grouping in a modified Ternus display. *Vision Research*, *47*, 2353–2366.

Wandell, B. A. (1999). Computational neuroimaging of human visual cortex. *Annual Review of Neuroscience*, *22*, 145–173.

Warrier, C. M., & Zatorre, R. J. (2004). Right temporal cortex is critical for utilization of melodic contextual cues in a pitch constancy task. *Brain*, *127*(Pt 7), 1616–1625.

Watson, J. D. G., Myers, R., Frackowiak, R. S. J., Hajnal, J. V, Woods, R. P., Mazziotta, J. C., … Zeki, S. (1993). Area V5 of the human brain: evidence from a combined study using positron emission tomography and magnetic resonance imaging. *Cerebral Cortex*, *3*(2), 79–94.

Werner, H. (1925). Über mikromelodik und mikroharmonik. *Zeitschrift Für Psychologie*, *98*, 74–89.

Wertheimer, M. (1912). Experimentelle Studien über das Sehen von Bewegung [Experimental studies on the seeing of motion]. *Zeitschrift Für Psychologie*, *61*, 161–265.

Wertheimer, M. (1922). Untersuchungen zur Lehre von der Gestalt, I: Prinzipielle Bemerkungen [Investigations in Gestalt theory: I. The general theoretical situation]. *Psychologische Forschung*, *1*, 47–58.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt, II [Investigations in Gestalt theory: II. Laws of organisation in perceptual forms]. *Psycologische Forschung*, *4*, 301–350.

Wessinger, C. M., VanMeter, J., Tian, B., Van Lare, J., Pekar, J., & Rauschecker, J. P. (2001). Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *Journal of Cognitive Neuroscience*, *13*(1), 1–7.

Weyl, H. (1952). *Symmetry*. Princeton, NJ: Princeton University Press.

White, B. W. (1960). Recognition of distorted melodies. *The American Journal of Psychology*, *73*(1), 100–107.

White, G. D. (2005). Musical scales and the tuning of musical instruments. In *Audio Dictionary* (3rd ed., pp. 495–499). Seattle, WA, USA: University of Washington Press.

Wickens, T. D. (2001). *Elementary Signal Detection Theory*. USA: Oxford University Press.

Widmann, A., Kujala, T., Tervaniemi, M., Kujala, A., & Schröger, E. (2004). From symbols to sounds: visual symbolic information activates sound representations. *Psychophysiology*, *41*(5), 709–715.

Williamson, V. J., Baddeley, A. D., & Hitch, G. J. (2010). Musicians' and nonmusicians' short-term memory for verbal and musical sequences: comparing phonological similarity and pitch proximity. *Memory & Cognition*, *38*(2), 163–175.

Williamson, V. J., Cocchini, G., & Stewart, L. (2011). The relationship between pitch and space in congenital amusia. *Brain and Cognition*, *76*(1), 70–76.

Wong, P. C. M., Ciocca, V., Chan, A. H. D., Ha, L. Y. Y., Tan, L.-H., & Peretz, I. (2012). Effects of culture on musical pitch perception. *PloS ONE*, *7*(4), e33424.

Yost, W. A. (2009). Pitch perception. *Attention, Perception, & Psychophysics*, *71*(8), 1701–1715.

Zacks, J. M. (2008). Neuroimaging studies of mental rotation: a meta-analysis and review. *Journal of Cognitive Neuroscience*, *20*(1), 1–19.

Zacks, J. M., & Michelon, P. (2005). Transformations of visuospatial images. *Behavioral and Cognitive Neuroscience Reviews*, *4*(2), 96–118.

Zatorre, R. J. (1988). Pitch perception of complex tones and human temporal-lobe function. *Journal of the Acoustical Society of America*, *84*(2), 566–572.

Zatorre, R. J., Bouffard, M., Ahad, P., & Belin, P. (2002). Where is "where" in the human auditory cortex? *Nature Neuroscience*, *5*(9), 905–909.

Zatorre, R. J., & Halpern, A. R. (1993). Effect of unilateral temporal-lobe excision on perception and imagery of songs. *Neuropsychologia*, *31*(3), 221–232.

Zatorre, R. J., Halpern, A. R., & Bouffard, M. (2010). Mental reversal of imagined melodies: a role for the posterior parietal cortex. *Journal of Cognitive Neuroscience*, *22*(4), 775–789.

Zatorre, R. J., Halpern, A. R., Perry, D. W., Meyer, E., & Evans, A. C. (1996). Hearing in the mind's ear: A PET investigation of music imagery and perception. *Journal of Cognitive Neuroscience*, *8*(1), 29–46.